Faculty Working Papers

A BOX-JENKINS ANALYSIS OF THE ADVERTISING-SALES RELATIONSHIP

Richard M. Helmer and Johny K. Johansson

#215

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

October 29, 1974

A BOX-JENKINS ANALYSIS OF THE ADVERTISING-SALES RELATIONSHIP

Richard M. Helmer and Johny K. Johansson

#215

ABSTRACT

The Lydia Pinkham data used by Palda and by Clarke and McCann to evaluate the lagged effects of advertising are reanalyzed using a Box-Jenkins transfer function analysis. After a general overview of the technique, the steps in the analysis are described and the empirical results at each stage reported. The need for proper pre-whitening of the advertising series is stressed. Final results indicate that there are no substantial effects of advertising beyond the first year -- confirming Clarke and McCann's cross-spectral analysis.

I. INTRODUCTION

Any decision maker needs to know the effectiveness of each course of action he is considering. This is a basic step in the scientific method of making decisions. In the absence of a well developed theory, the decision maker is forced to either experiment or examine past experience with the help of statistical analysis. A somewhat novel type of such empirical procedures has been developed by two applied statisticians, George E. P. Box of the University of Wisconsin and Gwilym M. Jenkins of the University of Lancaster, U. K. Their "Box-Jenkins transfer function analysis" is applicable to some of the more common problems that advertisers and researchers alike have attempted to answer. For example, are there only current effects of advertising or lagged effects also? If there is dynamic response is it short lived or continuing? Is the greatest effect of advertising immediate or delayed?

One of the first and best known investigations into lagged effects of advertising is Palda's [7]. His analysis of the Lydia Pinkham data showed that the response was dynamic, long lived, and had the greatest effect immediately.

More recently Clarke and McCann have reanalyzed the Lydia Pinkham data by Frequency Domain Analysis, a type of cross spectral analysis [3]. They concluded that no advertising effects were significant for periods longer than one year when using annual data and that the maximum effect occurred during the second month after advertising.

For this paper we have reanalyzed the Lydia Pinkham data in the time domain using Box-Jenkins procedures to see if additional insights into the lagged effects of advertising can be generated. In addition, since the Box-Jenkins approach is relatively new it might be of interest to see it applied to well known and readily available data.

The next section of this paper explains in greater detail the family of advertising effectiveness models that are considered in a Box-Jenkins transfer function analysis. Then a brief overview of the analytical stages in a Box-Jenkins approach is given. After that, the results of each stage in the actual analysis of the Pinkham data will be described and analyzed. After the discussion of the final stage results the conclusions with respect to the dynamic advertising effects are summarized. An appendix follows dealing in more detail with the statistics employed in the analysis.

II. THE BOX-JENKINS TRANSFER FUNCTION MODELS

1. The General Form

Box and Jenkins propose a rich set of response models as a family of transfer functions (also called impulse response functions). In their most general form the set of models can be written as the following discrete linear process

(1) $S_t = v_0 A_t + v_1 A_{t-1} + \cdots + N_t$

S_t = sales at time t
A_t = advertising at time t
N_t = sum of effects of all other variables other than advertising.

As a matter of notational convenience we employ a <u>backshift</u> operator B which is defined as

(2) $BZ_t = Z_{t-1}$ or $B^m Z_t = Z_{t-m}$.

We can therefore rewrite (1) as

$$(3)\quad S_t = (v_o + v_1 B + v_2 B^2 + \cdots)A_t + N_t$$
$$= v(B)A_t + N_t.$$

The polynomial operator v(B) is defined as the transfer function relating sales to advertising and it summarizes the dynamic structure of the effect transferred from the advertising sequence to the sales sequence. A restriction on the v's is that if advertising is held at a fixed level A_o, then sales should eventually reach an equilibrium level Y_o (a stationarity requirement). This assumption often necessitates the differencing of the sales and advertising series to eliminate trends and other sources of non-stationarity.

A further discussion of stationarity and differencing is given with respect to the Pinkham data in the pre-whitening section presented below. Apart from this restriction imposed by stationarity, the transfer function can take any polynomial form. Thus, a vast many alternative lagged effects can be accommodated. The "Box-Jenkins" analysis basically consists of procedures for assessing which of the many alternative over-time responses is in fact the correct one. Given the transfer function it is at least conceptually possible to select $X_t,\ldots,X_{t+a}$ so as to achieve any desired $Y_t,\ldots,Y_{t+a}$. This is the subject of much of control theory. The substantial difficulties in formulating and deriving the optimal control solu-

tions will not be discussed here (the interested reader is referred to Aoki [1]). The present problem is a less normative one: how do we model response of sales to advertising irrespective of the control situation facing us?[1]

2. The Polynomial v(B)

In the Box-Jenkins analysis the general polynomial v(B) is represented by the ratio of two polynomials of small degree compared to the degree of v(B). For example, if there were small response coefficients up to lag 3, after which there was a geometric decay in the coefficients we could express these coefficients in the following polynomial:

$$B^3 + \delta B^4 + \delta^2 B^5 + \ldots = \frac{B^3}{1-\delta B}$$

or (4) $$v(B) = \frac{B^3}{1-\delta B}$$

where δ is the decay coefficient. The general form of the transfer function is

(5) $$v(B) = \frac{\omega(B)}{\delta(B)} B^b$$

where $\omega(B)$ is a s^{th} order polynomial called the moving average operator;

$\delta(B)$ is a r^{th} order polynomial operator called the autoregressive operator;

B^b is a b^{th} order dead time operator;

---

[1] Strictly speaking, the situation facing the decision maker should influence choice of estimates and other statistical decisions (see Marschak [5]). In the present context we disregard this complication, however.

and r,s,b are integers greater than or equal to zero. Thus, the lag coefficients are specified once we have the polynomial. To get the polynomial, the Box-Jenkins analysis first derives the appropriate values of r, s, and b -- this is the "identification" problem, which uses cross-correlations (see Appendix). Given the values of these three identifying parameters, maximum likelihood estimates are then derived for the ω and δ parameters -- this is the "estimation" problem.

3. The Noise Process N_t

A further elaboration of the transfer function models accounts for the effect of situational and other unspecified factors called "noise." These factors, referred to as N_t, may be the composite effect of unaccounted for random shocks, past as well as present. The general form of the noise is

$$(6) \quad N_t = \frac{\theta(B)}{\phi(B)} (1-B)^d \, \varepsilon_t$$

where $\theta(B)$ is a q^{th} order moving average operator;

$\phi(B)$ is a p^{th} order autoregressive operator;

B^d is a d^{th} order difference operator;

ε_t is a Normal random variable;

and where p, d, and q are integers greater than or equal to zero.

As for the transfer function above, the parameters required to specify the noise process are derived in an identification and an estimation stage. The identification relies on autocorrelations and partial autocorrelations (see Appendix) to specify p, d, and q,

and then the maximum likelihood estimation of the θ and ϕ parameters is done conditional upon the assigned p, d, and q values.

This modeling of the noise process entails what is known as Box-Jenkins univariate analysis. This analysis, quite apart from its role in transfer function analysis, has been found useful in many applications (see, for example, Nelson [6]). In the transfer function analysis it is also used for the first stage of the analysis (the "pre-whitening" stage; see below) which builds a model of the input series specified by the parameters p, d, q, θ and ϕ .

III. THE TRANSFER FUNCTION MODEL OF THE LYDIA PINKHAM DATA

1. A Brief Overview

To clarify the structure of this part of the paper, a brief overview is in place. Generally, a transfer function analysis involves the choice of three models or processes. First, since the input series (advertising in our case) can be seen as strictly exogenous only if it is completely random, one transforms the given input data so as to achieve such randomness. This transformation involves the first choice of model: How should the input series be randomized or "pre-whitened"? The bases for the model choice are the autocorrelation and partial autocorrelation functions, and the empirical patterns are compared to alternative theoretical forms.

Second, after the output series (sales in our case) has been transformed similarly, the cross-correlation function between the transformed advertising and sales series is used to arrive at a

best representation of the effect of advertising upon sales. Again, empirical and theoretical patterns are compared for the model choice. Third, the residuals of the fit are looked upon as another time series, and a new univariate analysis (as in the first pre-whitening stage) is applied to derive a transformation of this "noise" process so as to make it completely random. The Box-Jenkins transfer function analysis consists of the derivation of a pre-whitening model, an impulse response function (the transfer function proper), and a noise model.

It should be emphasized that the approach is wholly empirical. The comparisons between various correlation patterns are based upon what a model with a particular parameter structure generates in terms of correlations, not upon a priori theory of, say, the advertising-sales relationship. Partly as a consequence of this empiricism, and as can be seen from the analysis to follow, many of the model choice questions will have to be resolved on fairly ambiguous bases. For any particular correlation pattern found in a sample, there will generally be many alternative models that could have generated the data. As a consequence, a good deal of the model choices become a matter of art. Because of this it becomes even more necessary than usual to present the statistical evidence upon which the choices and rejections were based.

Since the pre-whitening stage (which represents the first step in the analysis) contains many of the procedures followed in the other two stages, we will give a somewhat undue emphasis to that stage in our presentation. Also, the forecasting aspect is neg-

lected in this paper although this might often be the aim of a typical Box-Jenkins analysis. For example, the transfer function model might be used for forecasting, with the pre-whitening process first used to generate the future input values, and the impulse response function plus the noise process applied to generate the forecasts of the output series. In fact, the best test of the effectiveness of the advertising inputs might be to forecast sales on the basis of time alone -- using a univariate approach -- and then do the same with the transfer function model. If no improvement occurs with the latter, the advertising might possibly have no particular power to influence sales.

2. Pre-whitening of the Advertising Data

The aim of this section is to determine the appropriate form in which the advertising series should be used for the analysis in stages two and three. Ideally the advertising should have been randomly generated to avoid problems such as reverse causality. If the decision maker can engage in experimental randomization, for example, in a local test market, we would have perfect data with which to test the effect of changes in advertising. As it is, the first step in the analysis is to transform the actual advertising history to a random sequence. That is, we use transformations in order to eliminate the symptoms of nonrandomness in the independent variable. The signs of nonrandomness are generally heteroscedasticity of variance, lack of fixed mean, and autocorrelation.

A visual inspection of the raw advertising data in Figure 1(a) and (b) does not indicate h teroscedasticity. As can be seen, the later advertising values do not exhibit any particularly large swings as compared to the early ones.

Whether Figure 1 indicates a process with a fixed mean or not is more problematical. Both possibilities were entertained. The sequence of first differences of advertising clearly have a fixed (near zero) mean, as can be seen from Figure 2(a) and (b). However, we want to postpone the final choice of d=0, versus d=1, until we have considered the autocorrelations.

The autocorrelation at lag k, k=1,2,... is the correlation between realizations of the process separated by k periods (see Appendix). Significant autocorrelations in Figure 3 imply that there is a dependency between successive advertising inputs -- thus, advertising cannot be seen as randomly generated. As in the case of modeling the univariate noise process we eliminate this autocorrelation by identifying the p and the q of the autoregressive moving average process which transforms advertising to a random series.[1] We consider what kind of autocorrelation would be produced by each kind of possible p, d, q process and then compare this pattern of theoretical autocorrelations to the sample autocorrelations. Models with theoretical patterns distinctly different from the observed pattern are eliminated. Cases in which

---

[1]To preserve the relationship with the output (sales) series, the same transformation is applied to that series as well (see below).

NUMBER OF TERMS = 40 MEAN = 0.976575D 03 VARIANCE = 0.170645D 06

SERIES VALUES

1=	0.608000D 03	2=	0.451000D 03	3=	0.529000D 03	4=	0.543000D 03	5=	0.525000D 03
6=	0.549000D 03	7=	0.525000D 03	8=	0.578000D 03	9=	0.609000D 03	10=	0.504000D 03
11=	0.752000D 03	12=	0.613000D 03	13=	0.862000D 03	14=	0.866000D 03	15=	0.101600D 04
16=	0.136000D 04	17=	0.148200D 04	18=	0.160300D 04	19=	0.180000D 04	20=	0.194100D 04
21=	0.122900D 04	22=	0.137300D 04	23=	0.161100D 04	24=	0.156800D 04	25=	0.983000D 03
26=	0.104600D 04	27=	0.145300D 04	28=	0.150400D 04	29=	0.807000D 03	30=	0.339000D 03
31=	0.562000D 03	32=	0.745000D 03	33=	0.749000D 03	34=	0.862000D 03	35=	0.103400D 04
36=	0.105400D 04	37=	0.116400D 04	38=	0.110200D 04	39=	0.114500D 04	40=	0.101200D 04

Figure 1(a). Lydia Pinkham annual advertising expenditures from 1906 to 1935.

GRAPH OF SERIES

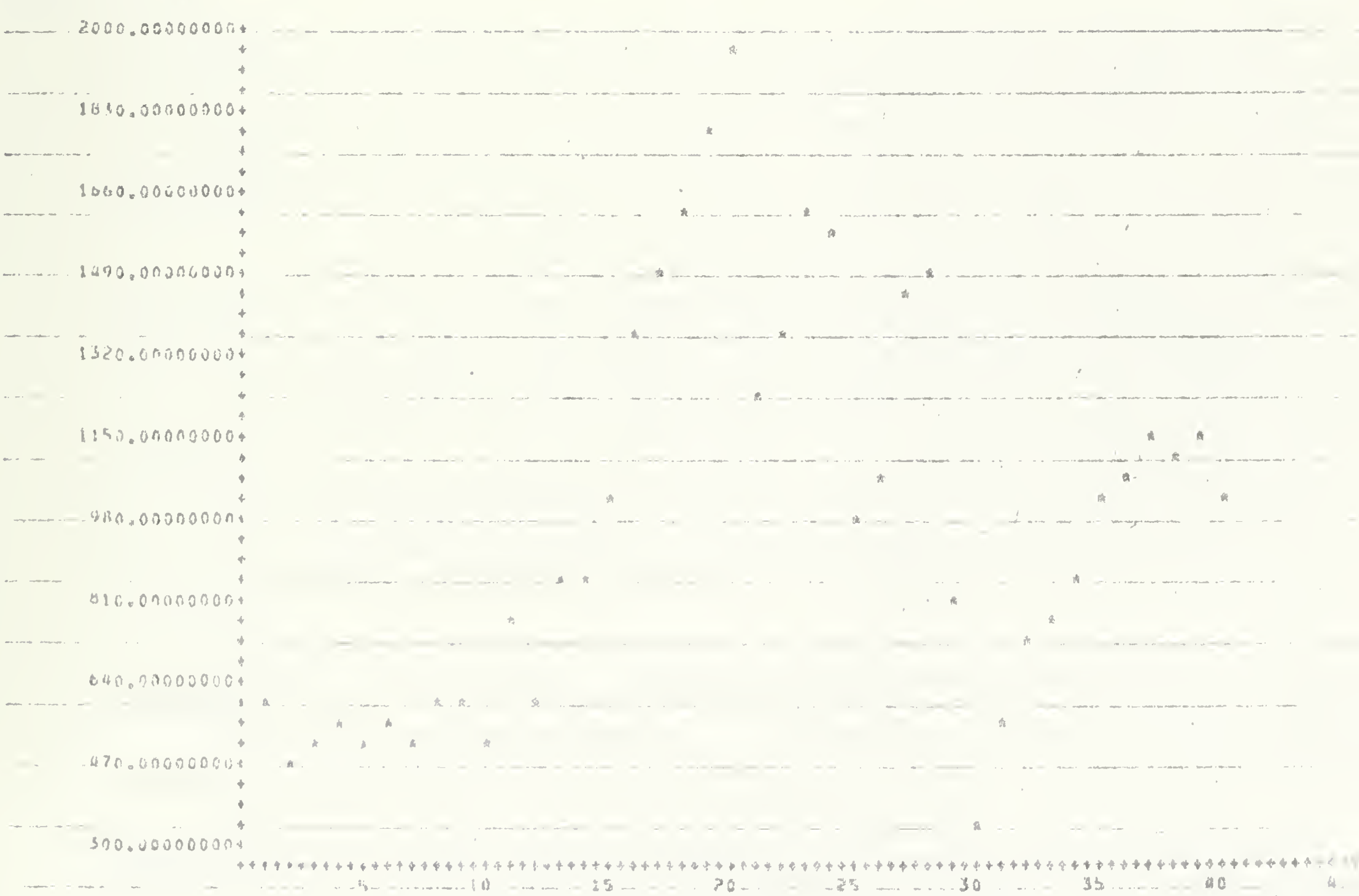


Figure 1(b). Graph of the Lydia Pinkham annual advertising expenditures.

NUMBER OF TERMS = 39 MEAN = 0.103590D 02 VARIANCE = 0.610543D 05

SERIES VALUES

1=	-0.157000D 03	2=	0.780000D 02	3=	0.140000D 02	4=	-0.180000D 02	5=	0.240000D 02
6=	-0.240000D 02	7=	0.530000D 02	8=	0.310000D 02	9=	-0.105000D 03	10=	0.208000D 03
11=	-0.139000D 03	12=	0.249000D 03	13=	0.400000D 01	14=	0.150000D 03	15=	0.344000D 03
16=	0.122000D 03	17=	0.126000D 03	18=	0.172000D 03	19=	0.141000D 03	20=	-0.712000D 03
21=	0.144000D 03	22=	0.238000D 03	23=	-0.430000D 02	24=	-0.585000D 03	25=	0.630000D 02
26=	0.407000D 03	27=	0.510000D 02	28=	-0.697000D 03	29=	-0.468000D 03	30=	0.225000D 03
31=	0.185000D 03	32=	0.100000D 03	33=	0.113000D 03	34=	0.172000D 03	35=	0.200000D 02
36=	0.110000D 03	37=	-0.620000D 02	38=	0.430000D 02	39=	-0.133000D 03		

Figure 2(a). First differences of advertising expenditures (advertising at year t minus advertising at year t-1).

GRAPH OF SERIES

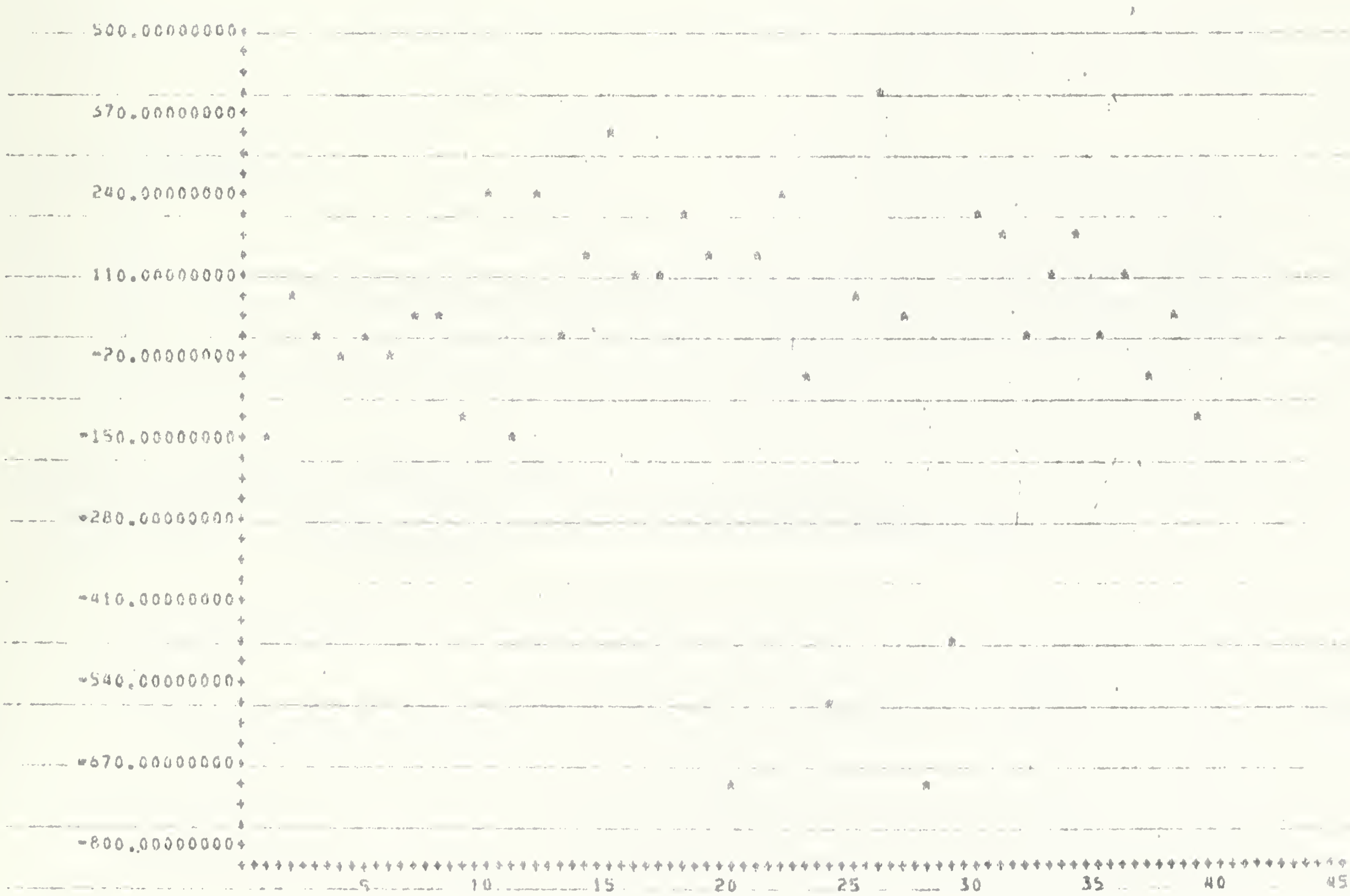


Figure 2(b). Graph of the first differences of advertising expenditures.

AUTOCORRELATION FUNCTION

LAG	0
VALUE	1.000

LAG	1	2	3	4	5	6	7	8	9	10	11	12
VALUE	0.615	0.600	0.528	0.495	0.309	0.108	0.023	-0.022	-0.126	-0.269	-0.330	-0.400

LAG	13	14	15	16	17	18	19	20
VALUE	-0.445	-0.452	-0.393	-0.337	-0.274	-0.204	-0.110	-0.050

APPROXIMATE STANDARD ERROR 1/SQRT(N) = 0.158

Figure 3(a). Autocorrelations of the advertising series.

AUTOCORRELATION FUNCTION

LAG	0
VALUE	1.000

LAG	1	2	3	4	5	6	7	8	9	10	11	12
VALUE	0.058	-0.400	-0.114	0.435	0.046	-0.337	-0.104	0.174	0.097	-0.195	-0.001	-0.045

LAG	13	14	15	16	17	18	19	20
VALUE	-0.115	-0.161	0.064	-0.012	-0.001	-0.034	0.123	0.033

APPROXIMATE STANDARD ERROR 1/SQRT(N) = 0.160

Figure 3(b). Autocorrelations of the first differences of advertising.

the empirical autocorrelations are somewhat similar to a theoretical pattern (allowing for sampling variations) are considered potential candidates for further investigation.

In addition to the autocorrelations, it is useful also to consider the partial autocorrelations before deciding upon the transformation (see Appendix). The partial autocorrelation function is a statistic which exploits the fact that autocorrelations at lag k may be a simple recursive function of autocorrelations at lags no greater than k. The sample partial autocorrelation at lag k is an estimate of the k^{th} autoregressive coefficient in the candidate process (Figure 4(a) and (b)).

Figure 5 shows the patterns of the autocorrelation function and the patterns of the partial autocorrelation function that can be derived for different p, d, q models. The graphs of the sample autocorrelations and partial autocorrelations of the advertising series are exhibited in Figures 6 through 9. The following p, d, q models can be considered most representative of the advertising process.

Alternative 1:	p=1, d=0, q=0 (1^{st} order autoregressive)
Supporting Evidence:	The autocorrelation function tails off (more linearly, however, than exponentially). (See Figure 6.) The partial autocorrelations have a major spike at lag 1 (see Figure 7).

LAG	1	2	3	4	5	6	7	8	9	10	11	12
VALUE	0.058	-0.404	-0.069	0.342	-0.098	-0.178	0.006	-0.117	0.043	-0.049	0.095	-0.234
LAG	13	14	15	16	17	18	19	20				
VALUE	-0.208	-0.151	-0.071	-0.109	0.132	-0.070	0.055	-0.065				

APPROXIMATE STANDARD ERROR 1/SQRT(N) = 0.160

Figure 4(a). The partial autocorrelations of the advertising expenditures.

LAG	1	2	3	4	5	6	7	8	9	10	11	12
VALUE	0.815	-0.193	0.314	-0.026	-0.411	-0.010	-0.004	-0.105	-0.010	-0.189	-0.003	-0.324
LAG	13	14	15	16	17	18	19	20				
VALUE	0.158	0.027	0.044	0.105	0.004	-0.081	0.089	-0.153				

APPROXIMATE STANDARD ERROR 1/SQRT(N) = 0.158

Figure 4(b). The partial autocorrelations of the first differences of advertising.

	Autoregressive processes	Moving Average processes	Mixed processes
Form	$A_t = \frac{1}{\phi(B)} \alpha_t$	$A_t = \theta(B)\alpha_t$	$A_t = \frac{\theta(B)}{\phi(B)} \alpha_t$
Autocorrelation function	infinite (damped exponentials and/or damped sine waves) tails off	finite cuts off	infinite (damped exponentials and/or damped sine waves after first $q - p$ lags) tails off
Partial autocorrelation function	finite cuts off	infinite (dominated by damped exponentials and/or sine waves) tails off	infinite (dominated by damped exponentials and/or sine waves after first $p - q$ lags) tails off

Figure 5. Patterns of the autocorrelation function and the partial autocorrelation function expected from autoregressive, moving average, and mixed autoregressive-moving average processes.

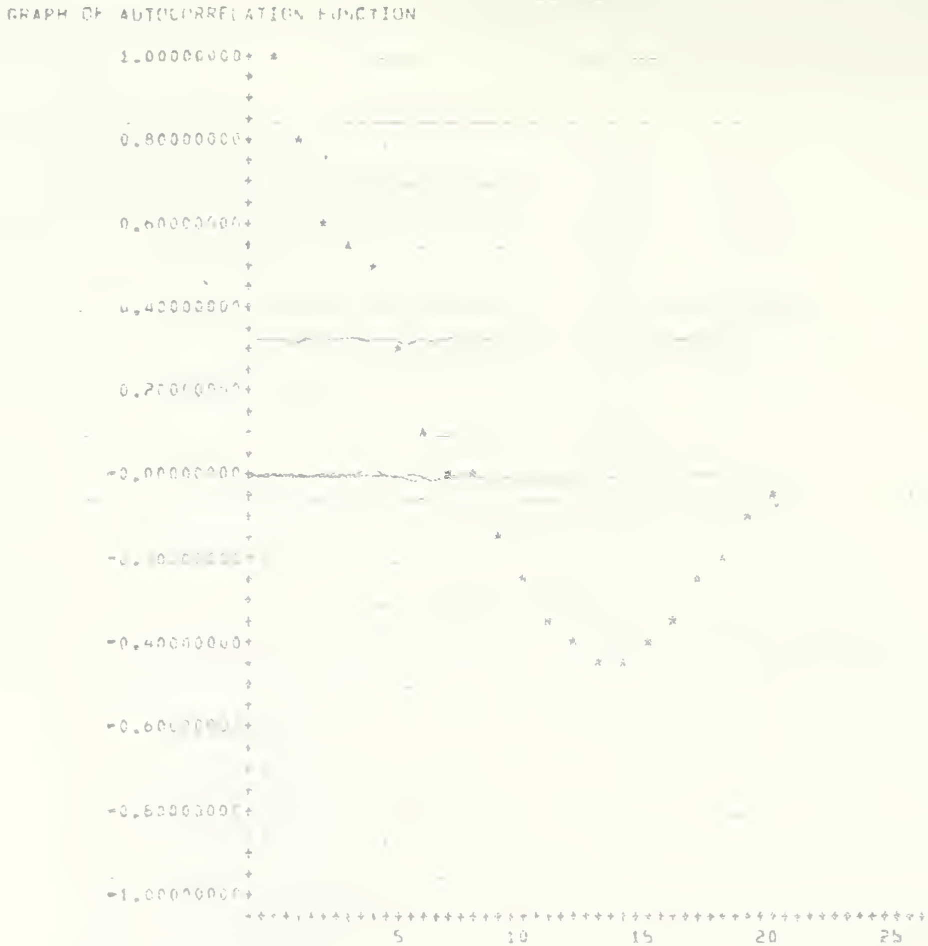


Figure 6. The autocorrelation function of advertising expenditures.

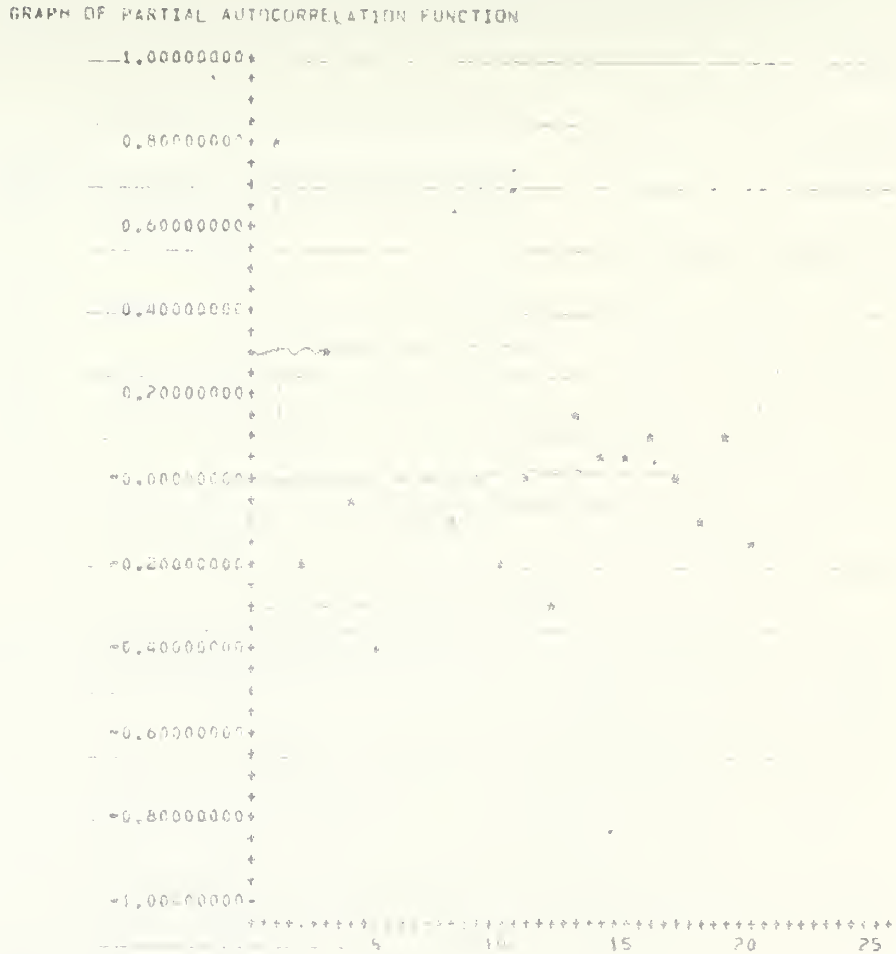


Figure 7. The partial autocorrelation function of advertising expenditures.

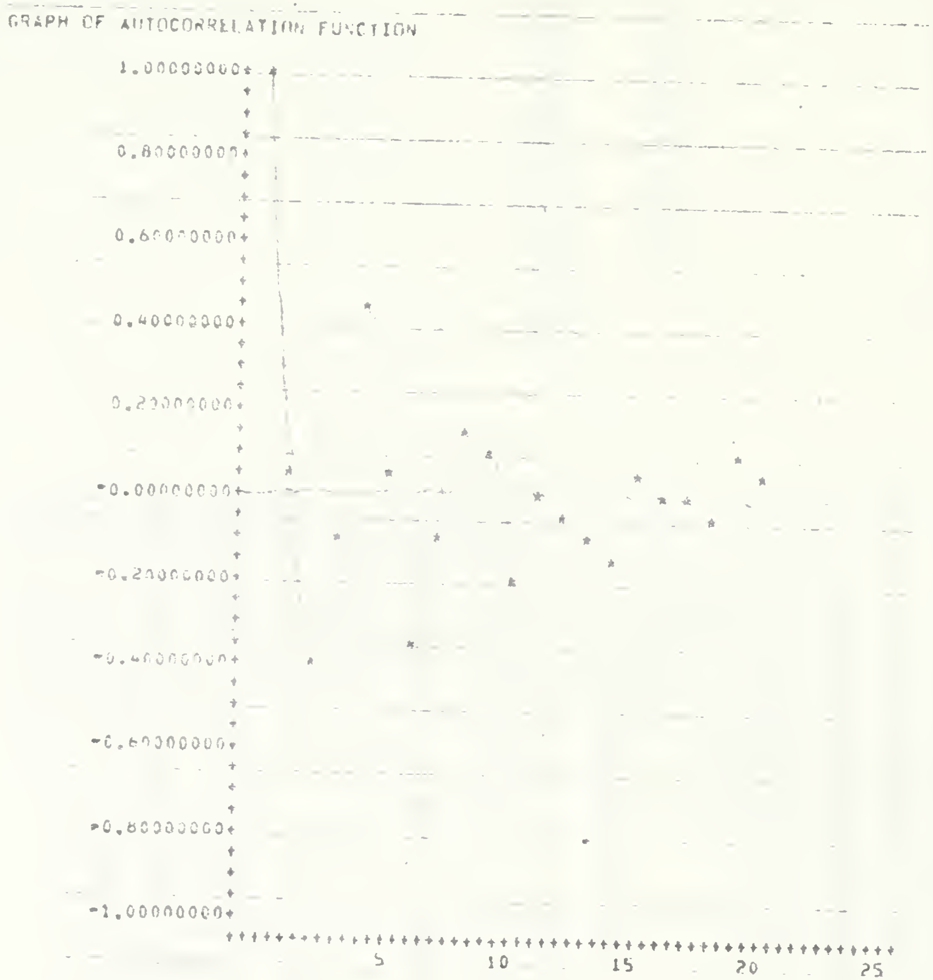


Figure 8. The autocorrelation function of first differences of advertising.

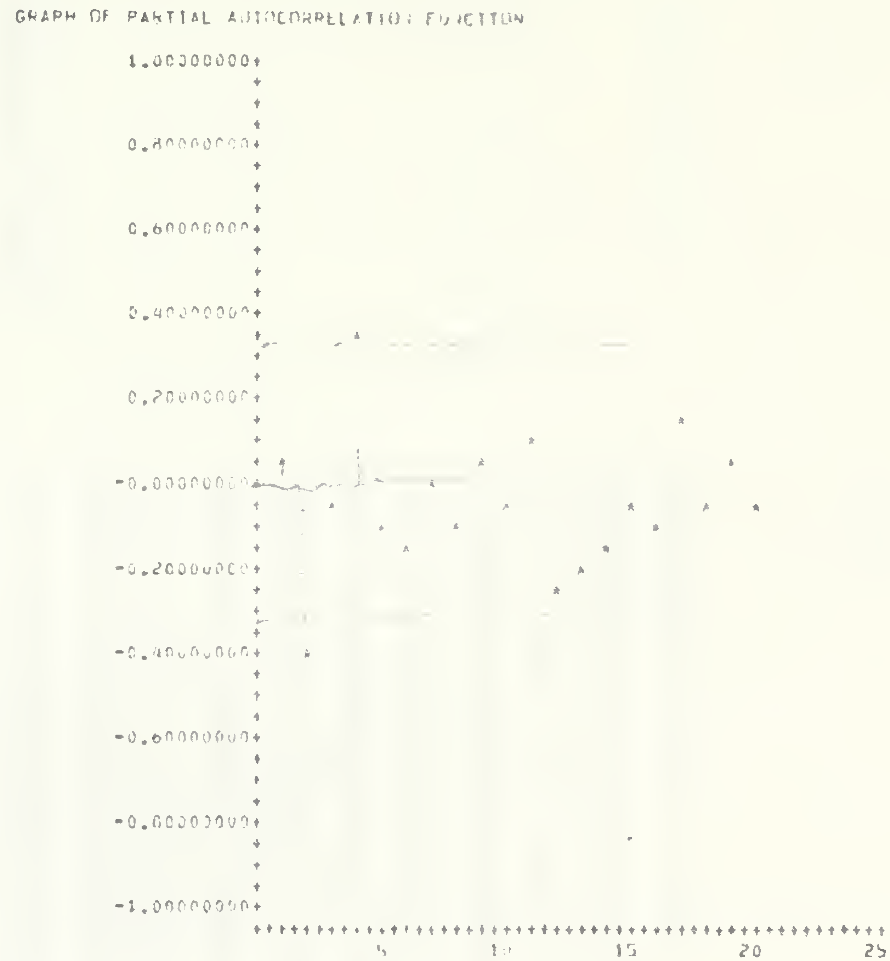


Figure 9. The partial autocorrelation function of first differences of advertising.

The spikes at lags 3 and 5 can be interpreted as caused by sampling error. Note: the approximate standard error of this estimate is for each partial autocorrelation taken separately and not the standard error of the function. Hence the larger the number of lags, the greater the possibility of partial autocorrelation exceeding 2 standard deviations.

<u>Refuting Evidence</u>: The relatively slow tapering off of the autocorrelations is an indication of possible non-stationarity (compare the models below where d=1).

<u>Alternative 2</u>: p=2, d=0, q=0
(2^{nd} order autoregressive)

<u>Supporting Evidence</u>: The full span of the autocorrelation function (see Figure 6) looks like a damped sine function. This pattern is expected from some (2,0,0) models. See this pattern in the lower right hand area of the triangle in Figure 10. The partial autocorrelation at lag 2 (see Figure 7) is negative as expected (see Figure 10).

Alternative 3:	p=2, d=1, q=0 (2nd order autoregressive, with 1st differencing)
Supporting Evidence:	The autocorrelation function (see Figure 8) has a distinct damped sine pattern indicative of the 2nd order autoregressive model (see Figure 10). The partial autocorrelations (Figure 9) are again first positive and second negative as expected (see Figure 10). The fourth autoregressive parameter is large and positive but discounted due to the strange behavior of those time periods, four years apart, occurring in 1926-27, 1930-31, and 1934-35.
Alternative 4:	p=1, d=1, q=1 (mixed 1st order autoregressive, moving average, with 1st differencing)
Supporting Evidence:	The partial autocorrelation (see Figure 9) looks like a damped sine function if one includes the positive spike at lag 4. This pattern can be explained by a mixed first order autoregressive and first order moving average process

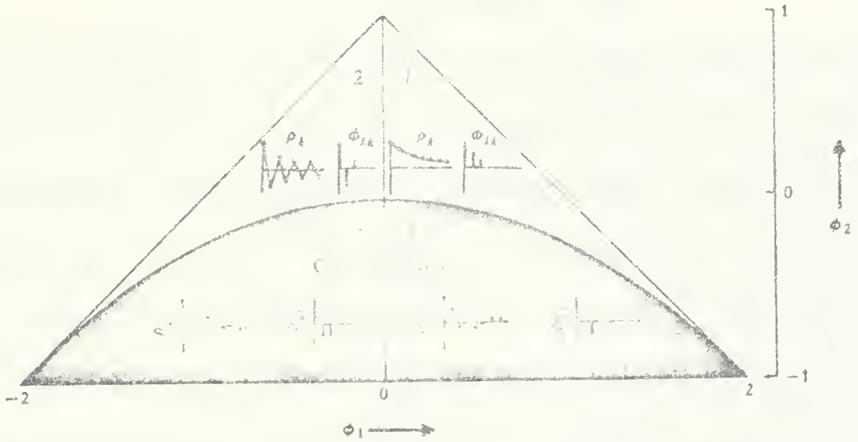


Figure 10. Typical autocorrelation and partial autocorrelation functions ρ_k and ϕ_{kk} for various second order autoregressive models.

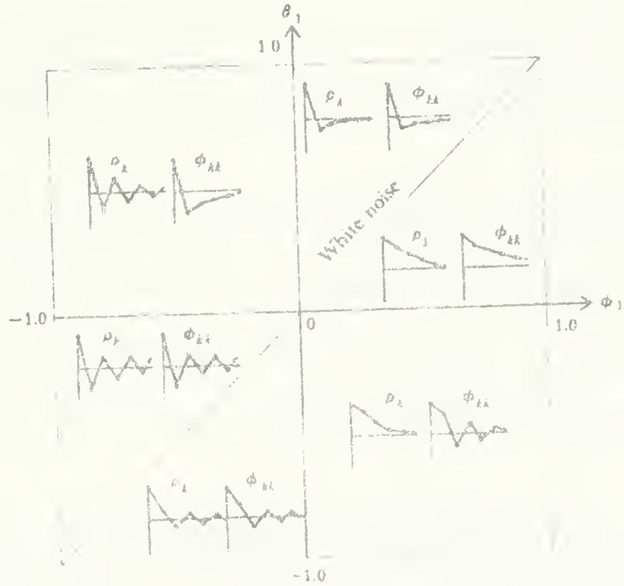


Figure 11. Typical autocorrelation and partial autocorrelation functions ρ_k and ϕ_{kk} for various mixed first order autoregressive - first order moving average models.

(see the lower right hand quadrant of Figure 11).

The final choice between these four specifications is not easy to make on the basis of the autocorrelations and partial autocorrelations alone, since, as we have seen all four versions can be supported. In such a case it is useful to proceed to the estimation of all four alternatives to get their respective $\hat{\theta}$ and $\hat{\phi}$ estimates and then make the choice of the basis of residual sum of squares, significance of the estimates, and other customary statistics.

The estimation of θ and ϕ is done using a maximum likelihood method with a non-linear least squares algorithm [2]. The estimates derived for the four models are presented in Figure 12. Judging from the standard errors of the estimates and the residual sum of squares, the specification p=2, d=1, and q=0 seems the best of the four. Consequently, this became our chosen pre-whitening transformation.[1]

If we have succeeded in transforming the advertising data to a random series then the series shown in Figure 13 should not be autocorrelated. Figures 14 and 15 show the autocorrelation and partial autocorrelation functions of the randomized series.

---

[1] The results in this section of the paper are based upon computer runs using only the first 40 observations. Similar -- in fact almost identical -- results were obtained when the full 54 years of data were used. In the transfer function analysis below, the prewhitening parameters $\hat{\theta}$ and $\hat{\phi}$ were based upon the full sample runs. The values estimated for the two parameters and then used below are $\hat{\theta}=0$, $\hat{\phi}_1 = .074$ and $\hat{\phi}_2 = -.407$.

(p,d,q) Model	Model with Maximum Likelihood Estimates (upper 95% Confidence limit) (lower 95% Confidence limit)	Residual Sum of Squares
(1,0,0)	$\alpha_t = A_t - \underset{\substack{(-.631)\\(-.999)}}{.815}A_{t-1} + \underset{\substack{(1453)\\(\ 610)}}{1032}$	$.21486\text{x}10^7$
(2,0,0)	$\alpha_t = A_t - \underset{\substack{(-.690)\\(-1.270)}}{.940}A_{t-1} + \underset{\substack{(\ .493)\\(-.161)}}{.166}A_{t-2} + \underset{\substack{(1390)\\(\ 690)}}{1040}$	$.20310\text{x}10^7$
(2,1,0)	$\alpha_t = (A_t - A_{t-1}) - \underset{\substack{(\ .216)\\(-.402)}}{.093}(A_{t-1} - A_{t-2}) + \underset{\substack{(.714)\\(.097)}}{.405}(A_{t-2} - A_{t-3})$	$.19566\text{x}10^7$
(1,1,1)	$\alpha_t = (A_t - A_{t-1}) - \underset{\substack{(\ .910)\\(-1.518)}}{.304}(A_{t-1} - A_{t-2}) + \underset{\substack{(1.161)\\(-.643)}}{.489}\alpha_{t-1}$	$.22722\text{x}10^7$

Figure 12. Maximum likelihood estimates, confidence intervals, and residual sum of squares of four candidate pre-whitening transformations of advertising.

PARAMETER VALUES VIA REGRESSION

1	2
0,9340D-01	-0,4055D 00

SUM OF SQUARES AFTER REGRESSION = 0,1956616D 07

ITERATION STOPS - RELATIVE CHANGE IN EACH PARAMETER LESS THAN 0,2000D-01

ANNUAL ADVERTISING OF LYDIA PINKHAM MEDICINE

GRAPH OF RESIDUALS

GRAPH INTERVAL IS 0,1618D 02

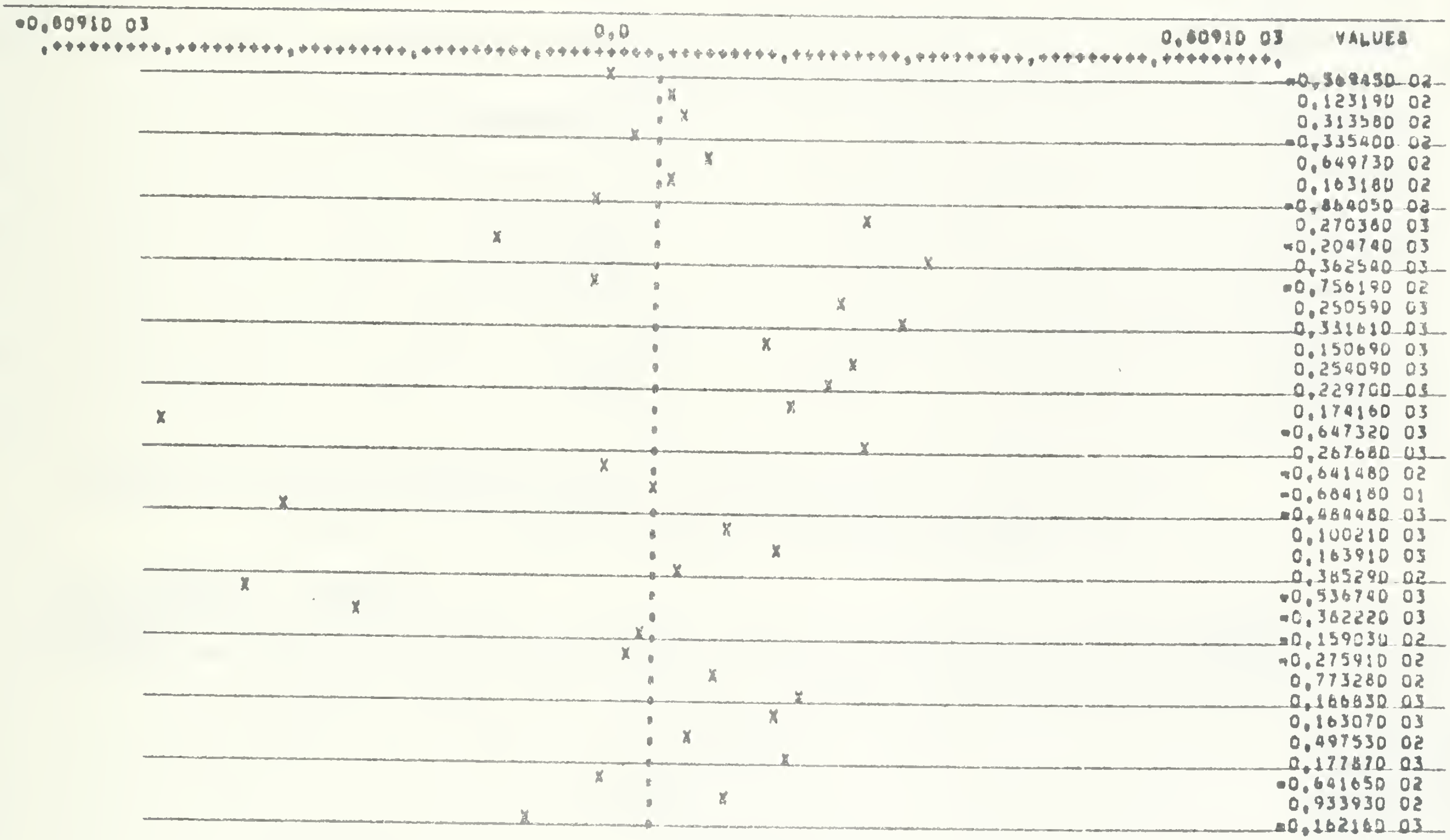


Figure 13. Residuals between the actual advertising expenditure series and the series of values generated by the (2,1,0) pre-whitening transformation.

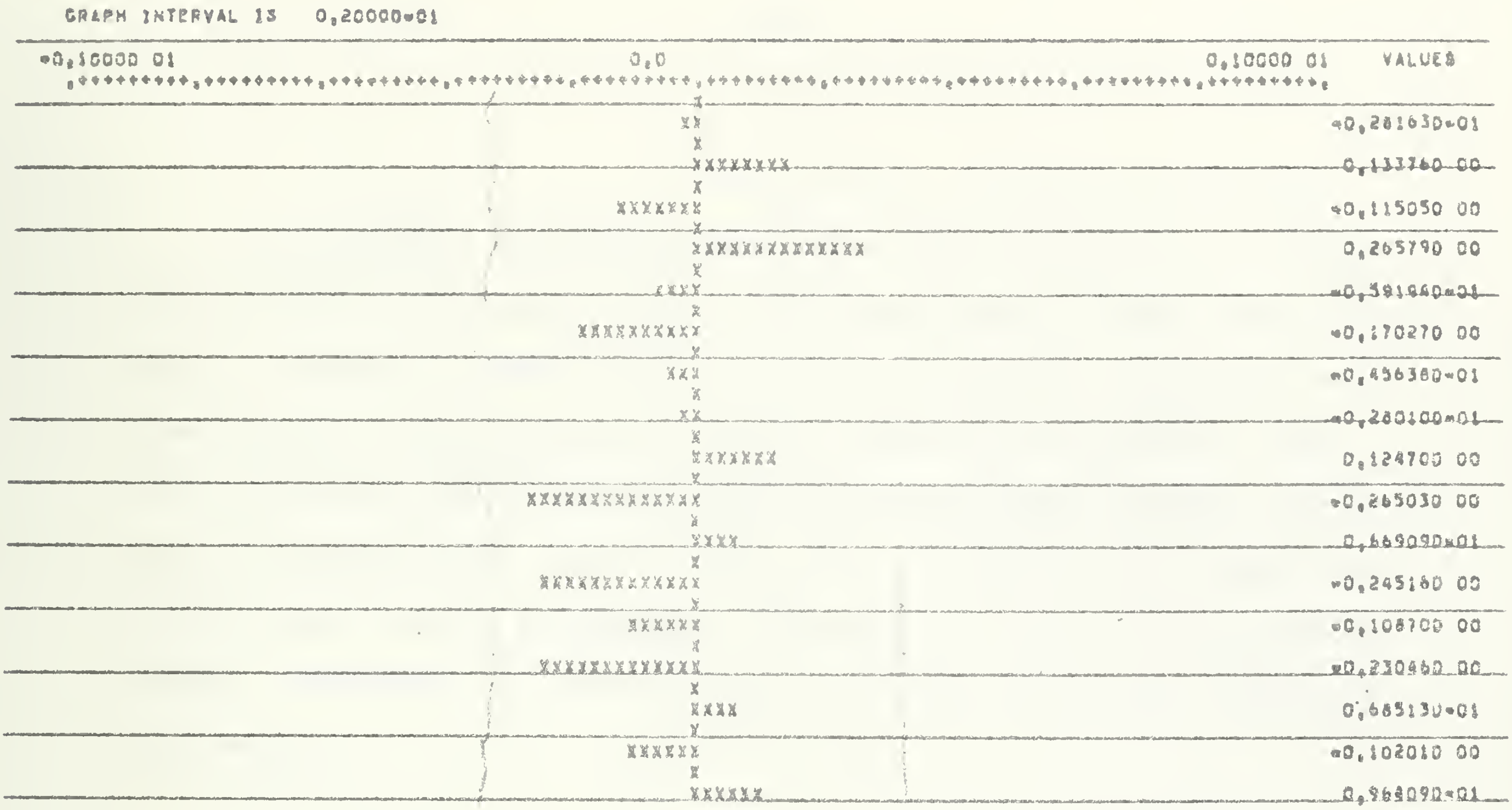


Figure 14. The autocorrelation function of the residuals of the (2,1,0) prewhitening transformation.

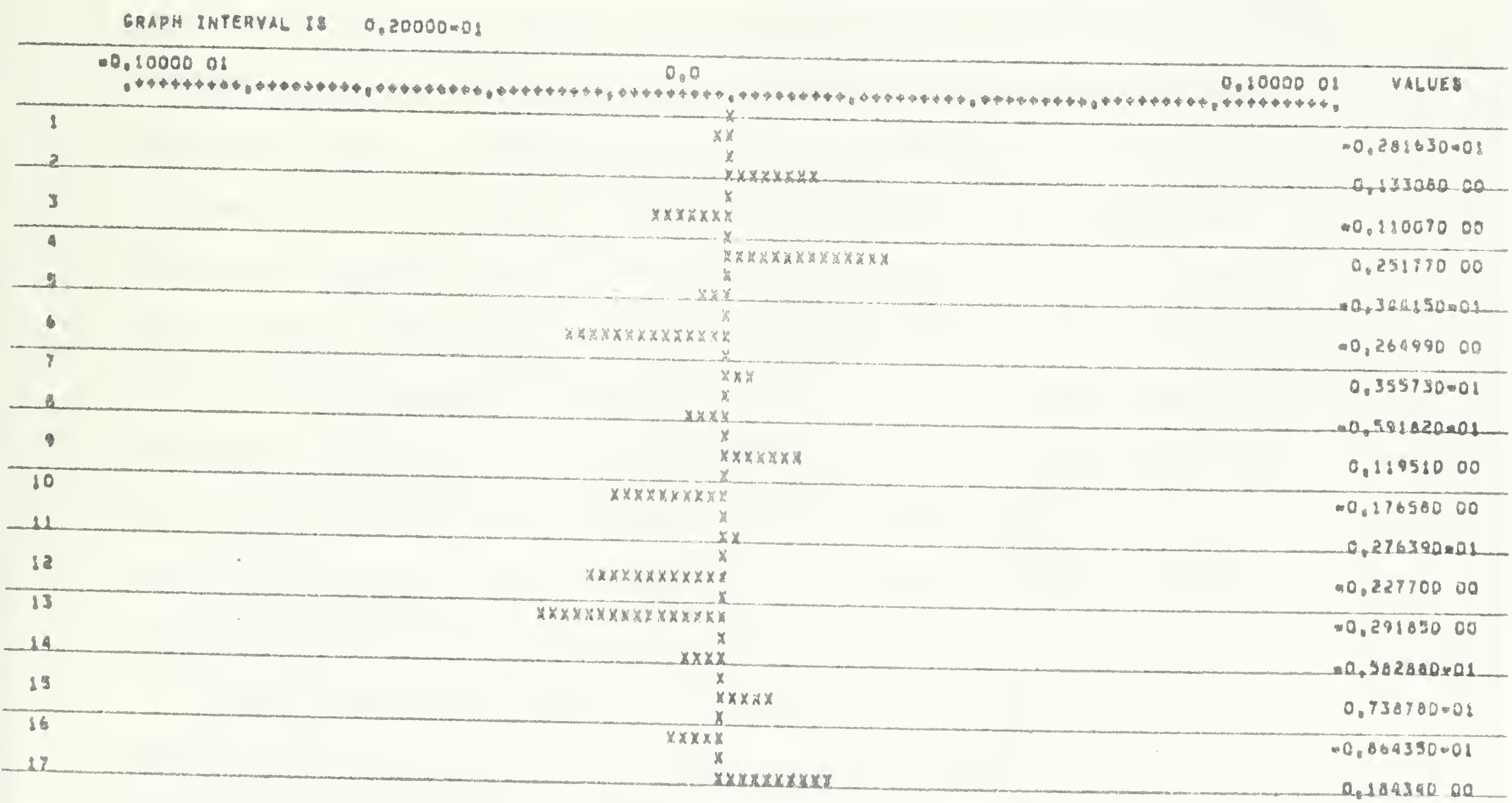


Figure 15. The partial autocorrelation function of the residuals of the (2,1,0) pre-whitening transformation.

A Chi-square statistic which tests the smallness of a whole set of sample autocorrelations can be used here to test for randomness. Figure 16 shows the value of the Chi-square statistic for the selected pre-whitening, and, for comparative purposes, also the Chi-square values of the other three candidate models. Clearly (2,1,0) emerges with substantially less autocorrelation than any of the other models considered. χ^2 around 8 is near the expected value. Therefore the autocorrelation is near the amount expected by chance.

3. Deriving the Impulse Response Function

On the basis of the pre-whitening function derived in the previous section, we will now proceed to the second stage of the analysis, the derivation of the impulse response (or transfer) function itself. Having gone into great detail in the previous discussion, we are now in a position to treat the procedural questions somewhat more superficially. Again, as we will see, the ultimate model choice is based upon rather artful considerations of correlational output, in this case the cross-correlation function. Another similarity with the previous analysis is the use of an identification stage -- here determining r, s, b -- and an estimation stage -- here of parameters ω and δ. These quantities were defined above in section II.

Even before the identification analysis begins, however, some transformations of the output variable (sales) will have to

(p,d,q) Model	Q statistic	degrees of freedom
(1,0,0)	18.7	8
(2,0,0)	18.5	7
(2,1,0)	8.3	8
(1,1,1)	19.2	8

Figure 16. Q statistics for candidate pre-whitening transformations. The Q statistic is distributed like a Chi-square.

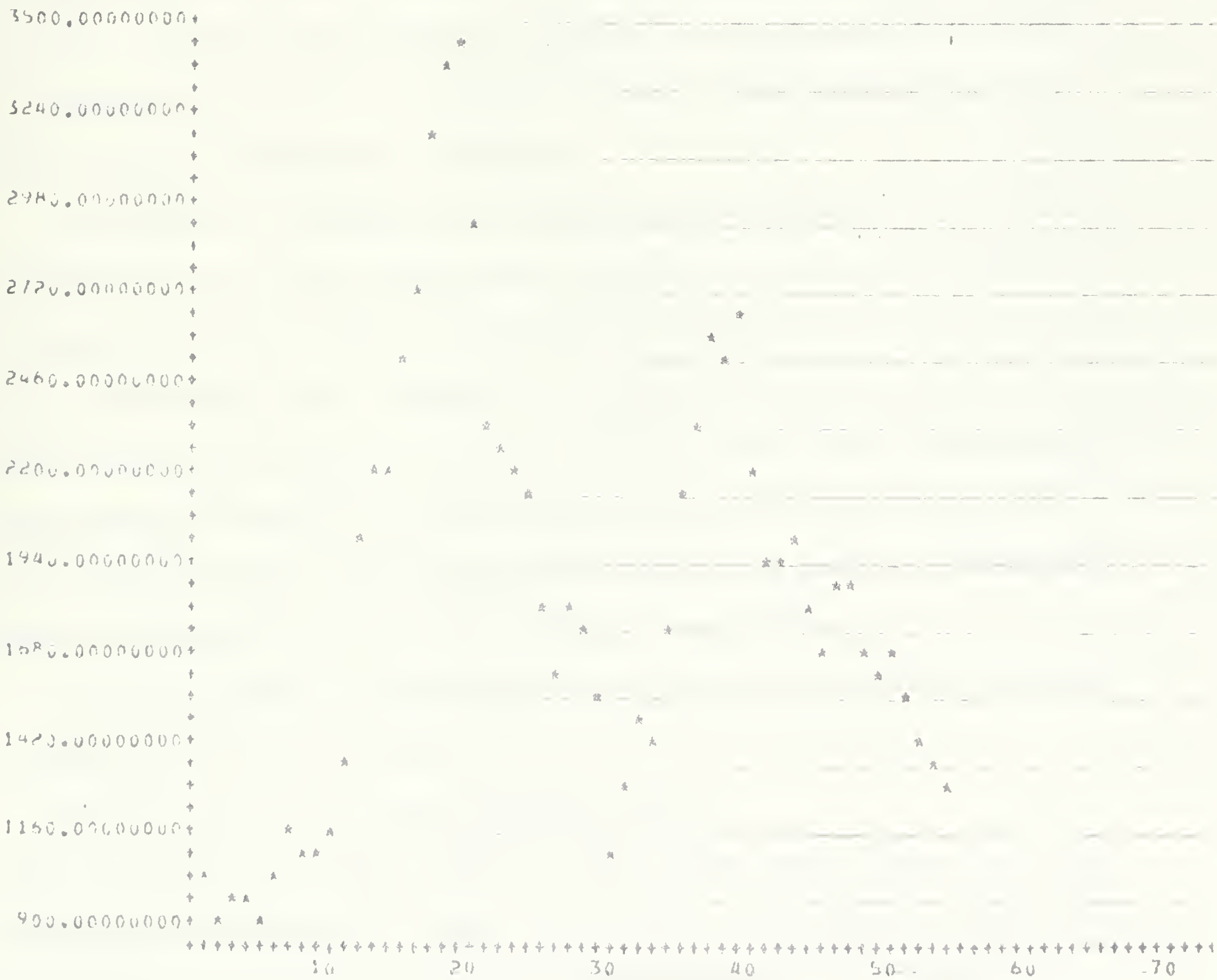


Figure 17. Graph of the Lydia Pinkham annual sales from from 1906 to 1960.

be made. First, as mentioned in the very beginning, there is a need for sales to be stationary so that some equilibrium level might be reached for a fixed advertising level. This might necessitate a differencing of the original sales series, identical to the differencing done in the univariate pre-whitening approach. The graphs of the original sales series and the series differenced once appear in Figures 17 and 18. As can be seen, the patterns are very much similar to the advertising data discussed earlier. Judging from these patterns and the autocorrelations in Figures 19 and 20, we decided to take first differences of sales for the subsequent analysis.

It should be pointed out that with the given pre-whitening transformation of the advertising series and with the present transformation of sales, the transfer function (1) of section II. will relate the first differences of both sales and advertising, rather than the original values. Second, since the pre-whitening transformation applied to the advertising data will affect the cross-correlations between the two series, we need to transform also the sales data with the same p, q operators.

In the preceding section we found the pre-whitening process to have p=2, d=1, q=0 and so

(7) $\quad \alpha_t = (1-.074+.407B^2)(1-B)X_t$

where X_t is the original advertising series and $(1-B)X_t = A_t$ is a differencing needed to make the inputs vary about a fixed mean. Applying the same pre-whitening transformation in (7) to the

THE SALES DATA

STAGE OF DIFFERENCING = 1 CORRESPONDING TO BASE SERIES DIFFERENCED WITH D = 1 DS = 0 S = 0

GRAPH OF SERIES

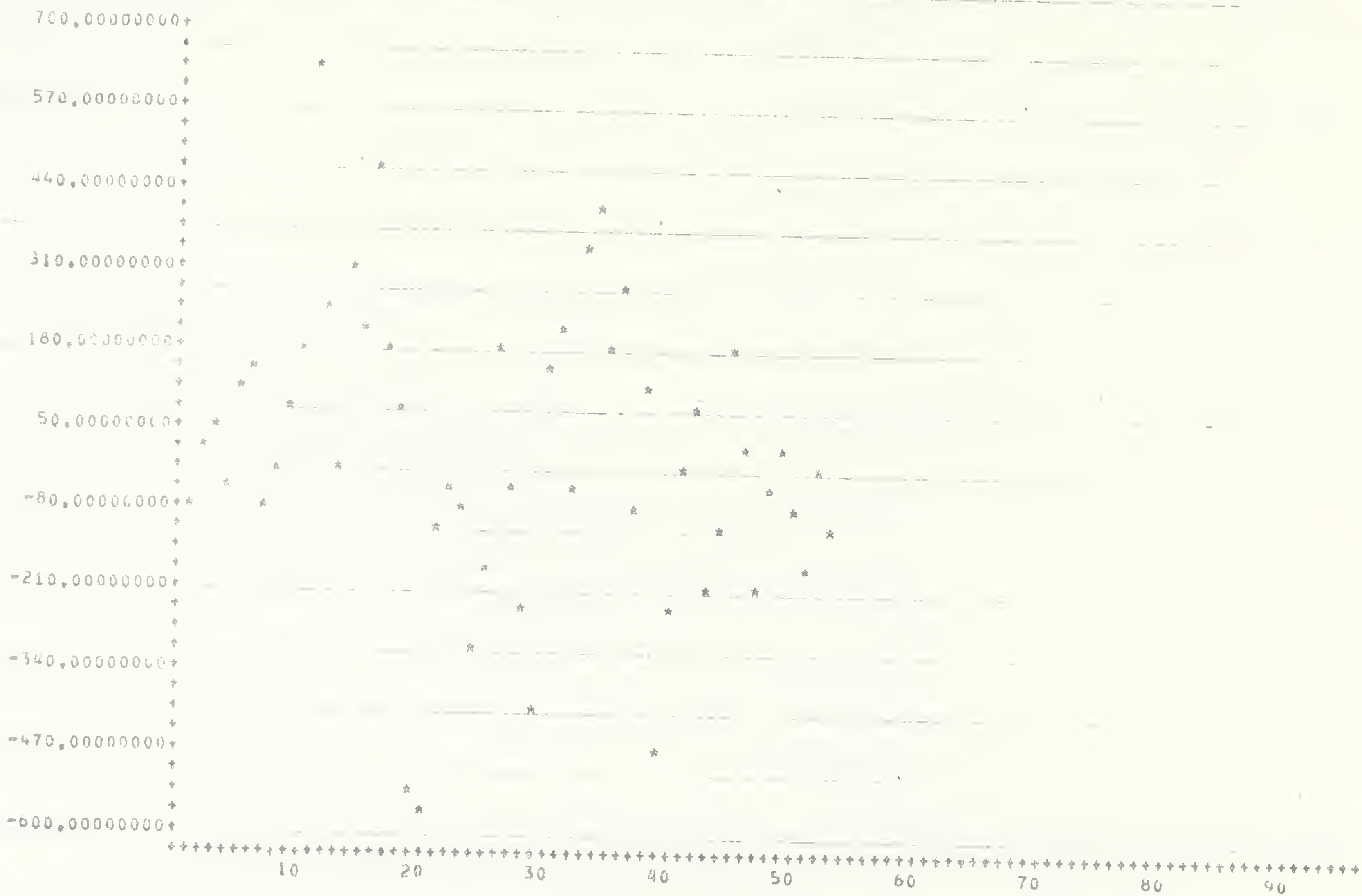


Figure 18. Graph of the first differences of the annual sales series.

THE SALES DATA

BASE SERIES WITHOUT DIFFERENCING

GRAPH OF AUTOCORRELATION FUNCTION

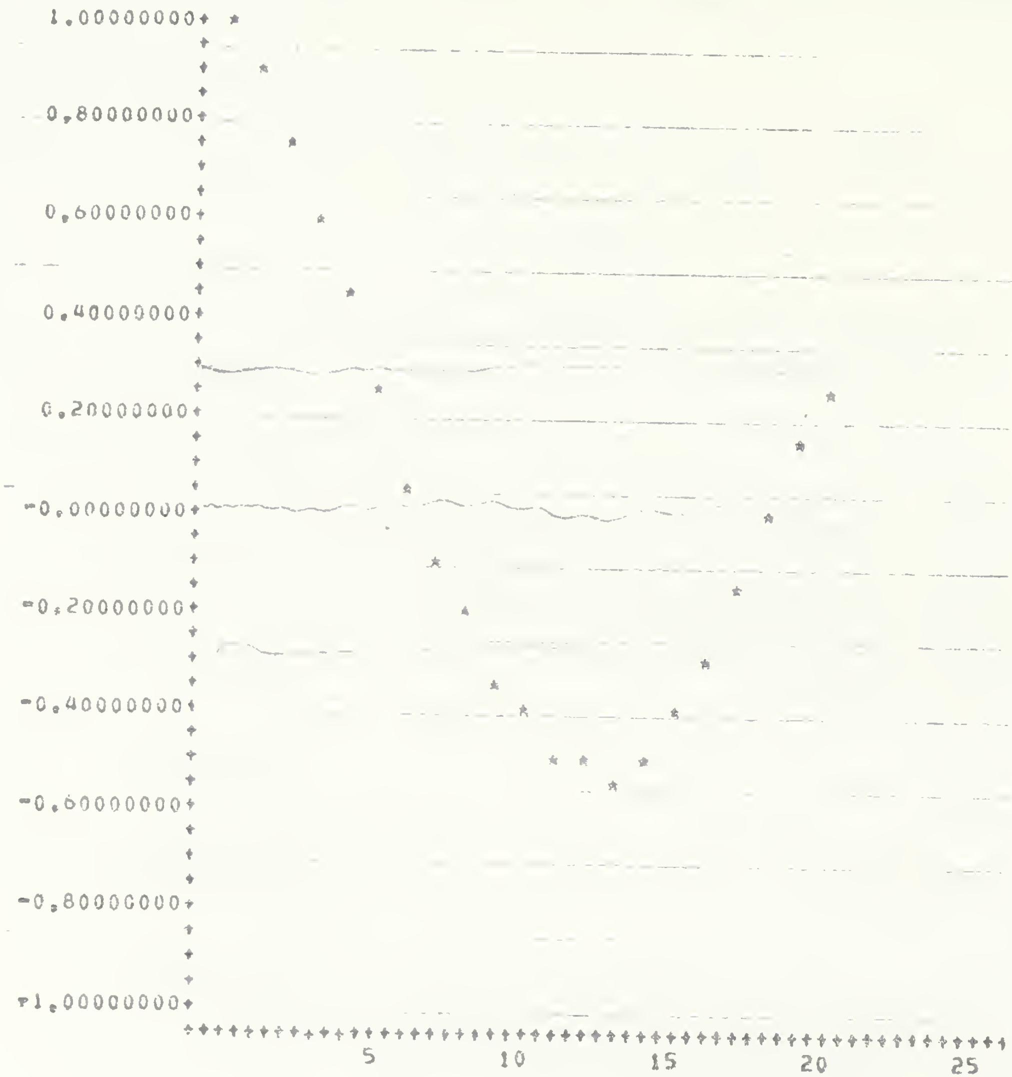


Figure 19(a). The autocorrelation function of the sales series.

GRAPH OF PARTIAL AUTOCORRELATION FUNCTION

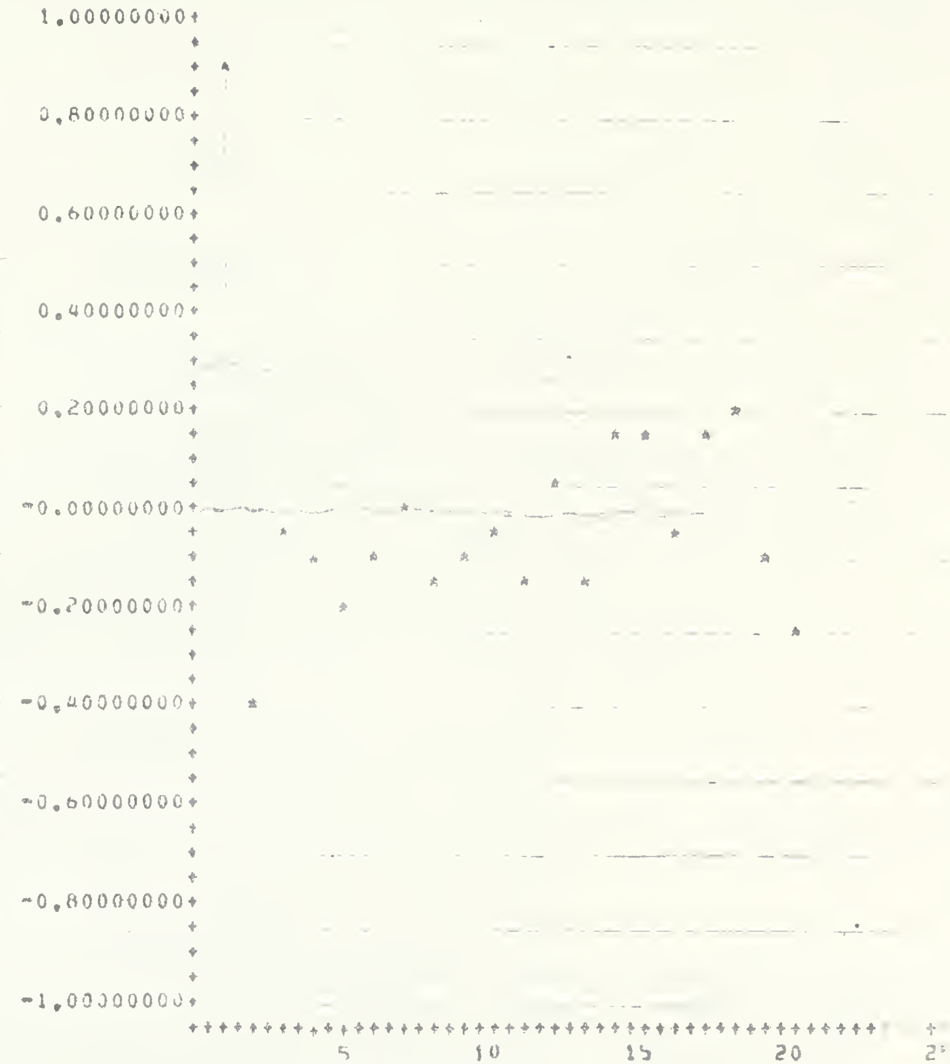


Figure 19(b). The partial autocorrelation function of the sales series.

THE SALES DATA

STAGE OF DIFFERENCING = 1 CORRESPONDING TO BASE SERIES DIFFERENCED WITH D = 1

GRAPH OF AUTOCORRELATION FUNCTION

Figure 20(a). The autocorrelation function of the first differences of sales.

GRAPH OF PARTIAL AUTOCORRELATION FUNCTION

Figure 20(b). The partial autocorrelation function of the first differences of sales.

original sales Y_t we get

(8) $\quad \beta_t = (1-.074+.007B^2)(1-B)Y_t.$

A third possible source of transformations of the sales data precedes in fact the two given. It might very well be that we want to impose some particular functional form (say, logarithmic) upon the sales-advertising relationship because of prior knowledge. This can be done in the same way as in the usual econometric analyses, i.e. by transforming the individual variables first and then relating their transformed values linearly. If such a transformation is deemed desirable, it should be carried out before any of the analysis discussed so far, including the pre-whitening of the advertising series. In the present case, no such transformation of the data was made.

After these transformations have been completed, the identification analysis of the impulse response function can be carried out. The requisite r, s, and b parameters can be chosen on the basis of the cross-correlation function (see Appendix) between the transformed sales and advertising series. To see why these cross-correlations are of such importance -- they do in fact determine the impulse response function directly -- the following derivation will be instructive.

Recall that $(1-B)Y_t = S_t$. We replace S_t by (3) of section II. and multiply by α_{t-k}. Taking the expected value, and recognizing that the noise N_t does not covary with X_t, we have

(9) $\quad E(\alpha_{t-k}\beta_t) = E(\alpha_{t-k}(v(B)\alpha_t).$

Now,

$$(10) \qquad E(\alpha_{t-k}\beta_t) = E(\alpha_{t-k}(v_o\alpha_t - v_1\alpha_{t-1} - v_2\alpha_{t-2}\cdots))$$

$$(11) \qquad = v_k E(\alpha_{t-k})^2$$

$$= v_k \sigma_\alpha^2$$

since α_t's are constructed to be uncorrelated shocks with zero mean and variance σ_α^2. The k^{th} impulse response coefficient is therefore simply related to the covariance of α_{t-k} and β_t, and thus to the cross-correlation

$$(12) \qquad v_k = \frac{\mathrm{cov}(\alpha_{t-k}\beta_t)}{\sigma_\alpha^2}$$

$$= \frac{\mathrm{corr}(\alpha_{t-k}\beta_t)\,\sigma_\beta}{\sigma_\alpha} .$$

Thus, the k^{th} coefficient of the transfer function is identical (except for a scale factor) to the cross-correlation between the transformed sales series and the pre-whitened advertising series lagged k periods.

The sample cross-correlation function of the sales-advertising relationship for the transformed Pinkham data is presented in Figure 21. On the basis of this graph the appropriate values of the r, s, and b parameters can be chosen (see Figure 22), just as in the case of the p, d, and q parameters of the univariate analysis. Comparing the theoretical patterns in the table with the sample correlations in the graph, we conclude that there are two candidate r, s, and b model specifications, corresponding to two alternative transfer functions.

LAG	0	1	2	3	4	5	6	7	8	9
VALUE	0,5156	0,3239	0,1205	-0,0494	0,0321	-0,2494	-0,2485	-0,2646	-0,1564	-0,0763

LAG	10
VALUE	-0,1807

TANDARD ERROR SET TO SQRT(1/.), = 0,140

Figure 21(a). Cross-correlation function between the pre-whitened advertising and sales series.

.AG	0	1	2	3	4	5	6	7	8	9
/ALUE	0.6495	0.4080	0.1518	-0.0623	0.0404	-0.3141	-0.3129	-0.3332	-0.1994	-0.0961

.AG	10
/ALUE	-0.2275

;ANDARD ERROR = 0.176

Figure 21(b). The impulse response function (the transfer function) between the prewhitened advertising and sales series.

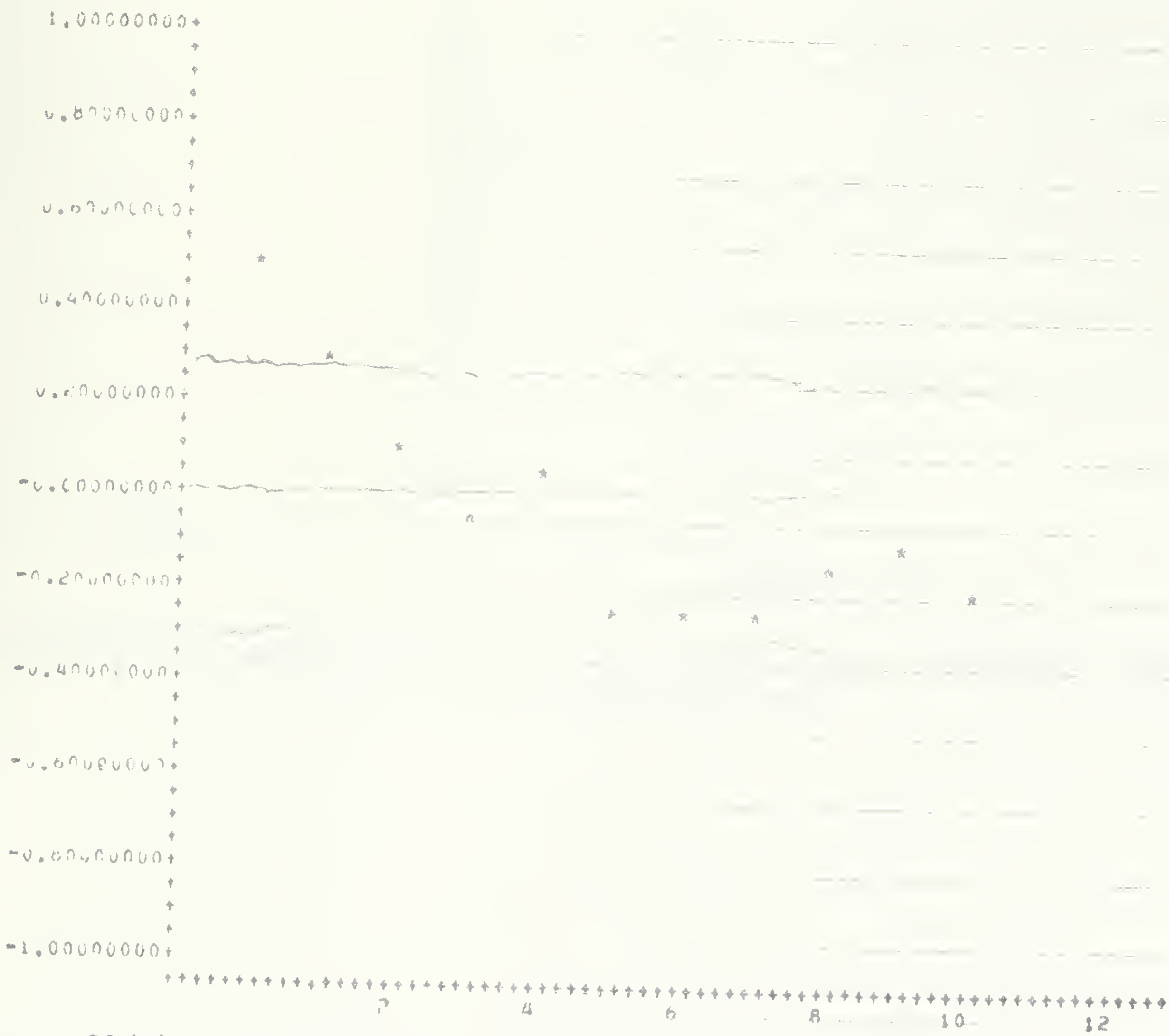


Figure 21(c). Graph of the cross-correlation function. The graph of the impulse response function is similar.

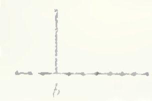
$r\,s\,b$	B Form	Impulse Response υ_j
003	$Y_t = B^3 X_t$	
013	$Y_t = (.5 + .5B) B^3 X_t$	
023	$Y_t = (.25 + .50B + .25B^2) B^3 X_t$	
103	$(1 - .5B) Y_t = .5B^3 X_t$	
113	$(1 - .5B) Y_t = (.25 + .25B) B^3 X_t$	
123	$(1 - .5B) Y_t = (.125 + .25B + .125B^2) B^3 X_t$	
203	$(1 - .6B + .4B^2) Y_t = .8B^3 X_t$	
213	$(1 - .6B + .4B^2) Y_t = (.4 + .4B) B^3 X_t$	
223	$(1 - .6B + .4B^2) Y_t = (.2 + .4B + .2B^2) B^3 X_t$	

Figure 22. Examples of impulse response functions (transfer functions) for various (r,s,b) specifications.

<u>Alternative 1</u>:	r=1, s=0, b=0 (1st order autoregressive)
<u>Supporting Evidence</u>:	The first three cross-correlations decline gradually, almost exponentially which would tend to argue for an autoregressive process. The later cross-correlations are irregular, some even negative, but none is significantly different from zero. This type of specification is a Koyck model and close to some of the Palda versions run. The major difference is that in our case both the advertising and the sales data have been differenced so as to achieve stationarity before being correlated.
<u>Alternative 2</u>:	r=0, s=1, b=0 (1st order moving average)
<u>Supporting Evidence</u>:	Taking notice of the fact that the cross-correlation at t-2 is, in fact, insignificant at the .05 level, one can argue that only the first two correlations should be considered. In addition, the cross-correlations at t-3 and t-4 are near zero so there

does not seem to be much of an autoregressive effect showing. This is close to the stand taken by Clarke and McCann in their spectral analysis. Again, there are differences due to the differing treatment of the stationarity requirement.

Because of the correspondence between the models here and the previous research, it is of special interest to see if the Box-Jenkins approach can resolve the conflict as to the number of periods over which advertising has substantial effects. Figure 23 is a graph of the respective lag coefficients for the two contending models and of the Box-Jenkins impulse response coefficients. The Box-Jenkins coefficients consist simply of the impulse response coefficients derived directly from the cross-correlations. In each case the lags after the first year are statistically insignificant at the .05 level. Note however that the values of the 0th and 1st lag coefficients are very close for all three models. The major discrepancy between the Palda and the Clarke and McCann models is in lags 2, 3, and 4 and higher which the Palda model shows as positive but the Clarke and McCann data show to be near zero. Our results show the 2nd to be positive while the 3rd and 4th are near zero. On the basis of the cross-correlations alone, nothing conclusive can clearly be said. At this stage it is obviously a matter of rather difficult judgment which one of the two processes should be chosen. A final choice is postponed until we obtain

	k	Palda's estimates based on $\delta = .628$ $\omega = .537$	Clarke-McCann estimates based on FDR	Box-Jenkins estimates based on cross-correlations
Coefficient at lag k	0	.537	.642	.649
	1	.337	.297	.408
	2	.211	.051	.152
	3	.133	.039	-.062
	4	.082	-.062	.040
	5	.052	-.189	-.314

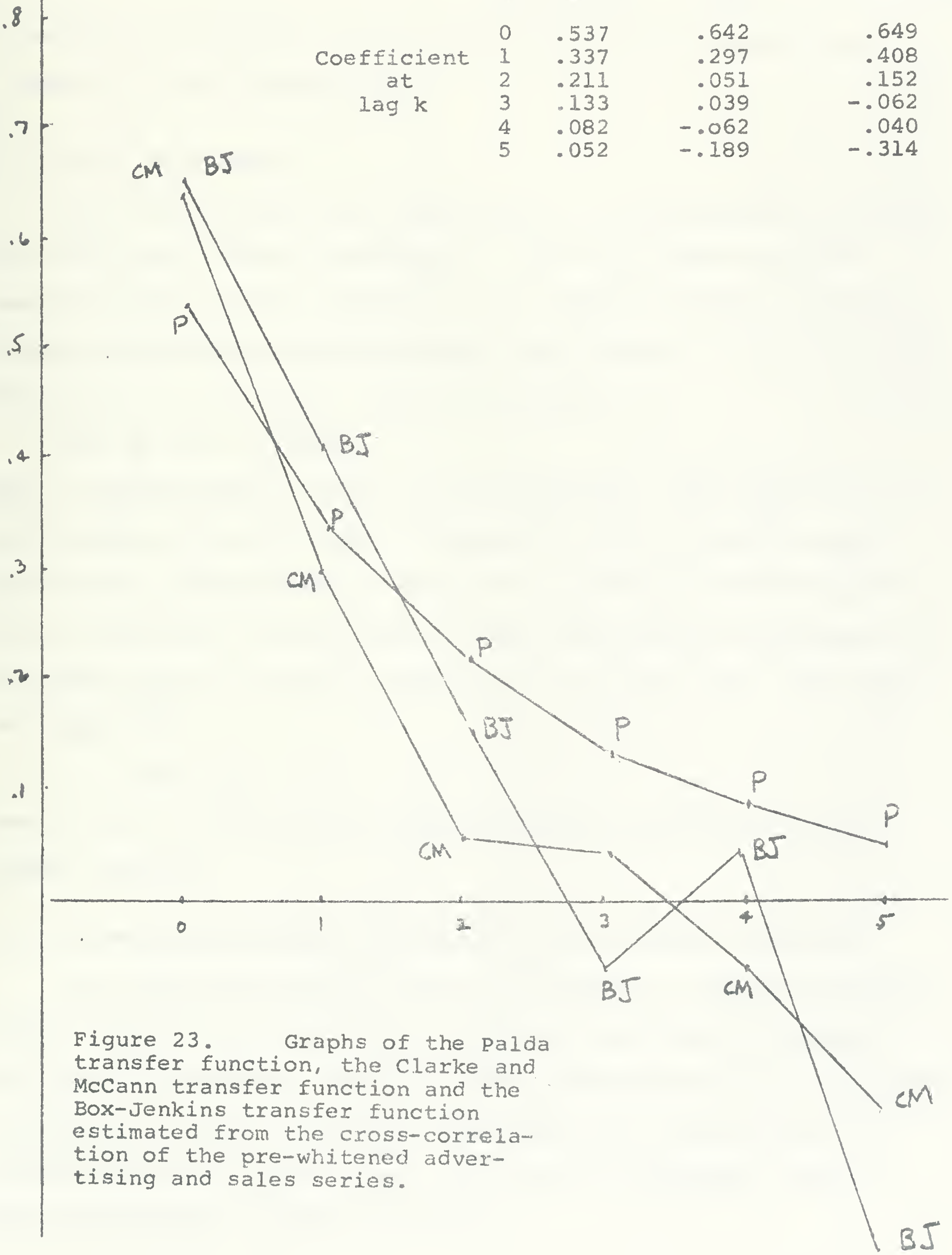


Figure 23. Graphs of the Palda transfer function, the Clarke and McCann transfer function and the Box-Jenkins transfer function estimated from the cross-correlation of the pre-whitened advertising and sales series.

preliminary estimates of the second part of the transfer function model, the noise process, N_t.

4. The Noise Process

The identification and estimation of the noise process follows the univariate analysis completely. The series under analysis consists simply of the residuals of the fitted transfer function. Without making any assumptions about the structure of this transfer function we can calculate residuals directly by

$$(13) \quad \hat{N}_t = Y_t - \hat{v}(B)X_t.$$

Here a limited order of $\hat{v}(B)$ of 10 is assumed. The residuals were quite random as can be seen in Figures 24 and 25. The Chi-square test of the autocorrelations showed no significant autocorrelation at the .05 level. On the basis of this identification procedure we conclude that p, d, and q in Equation (6) are all zero. Therefore, we simply have

$$N_t = \varepsilon_t$$

where ε_t is an independently and identically distributed Normal random variable.

5. The Maximum Likelihood Estimates of the Transfer Functions

As in the pre-whitening stage, we attempt to resolve the model choice question by calculating the maximum likelihood estimates of the ω and δ parameters, and then using goodness-of-fit and diagnostic statistics. Figure 26 shows the two models with the maximum likelihood estimates and the upper and lower 95% confidence limits of the two candidate transfer functions.

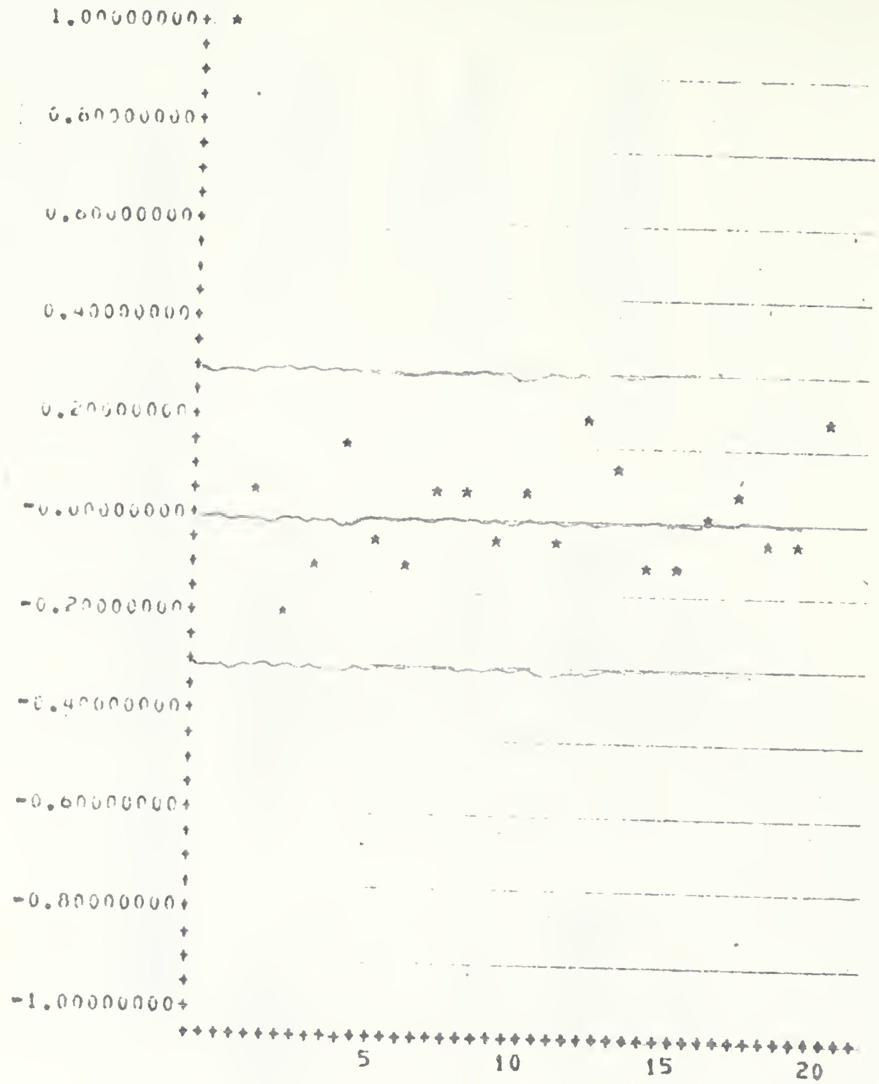


Figure 24. The Autocorrelation function of the noise process based on the cross-correlation function.

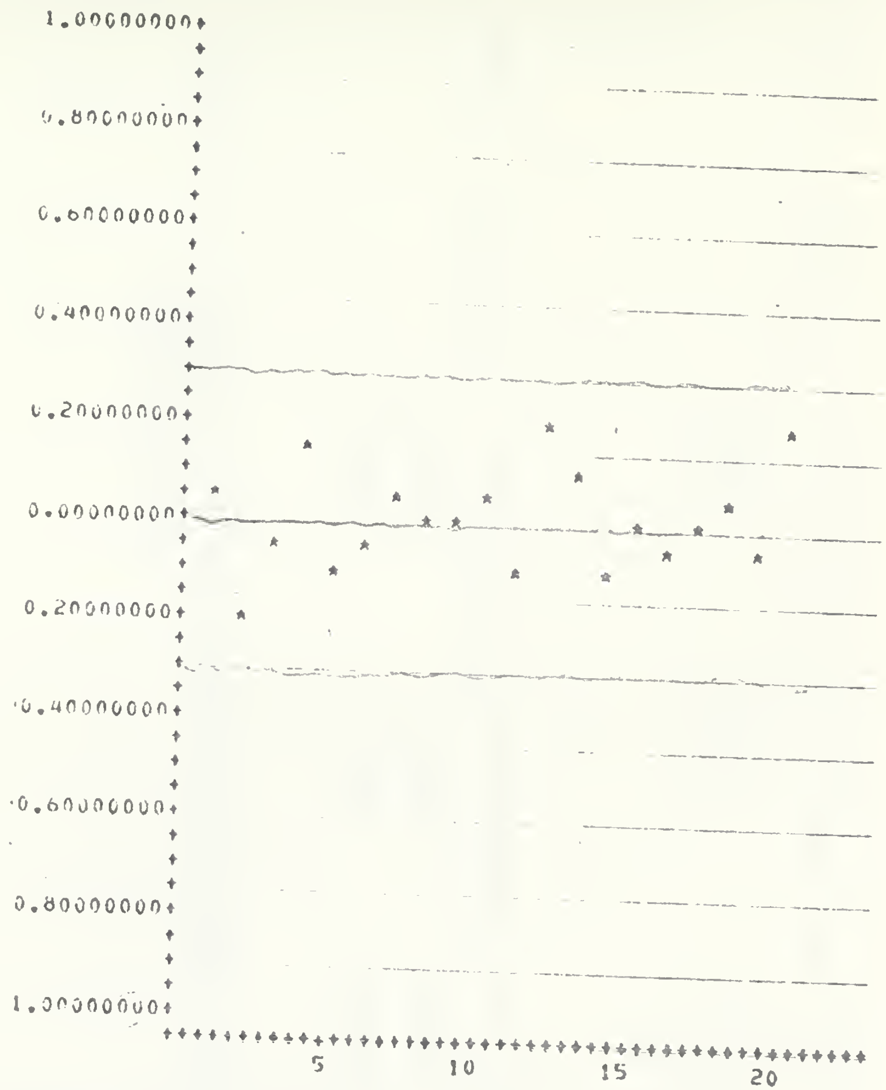


Figure 25. The partial autocorrelation function of the noise process based on the sample cross-correlation function between prewhitened advertising and sales.

(14) (010) $(Y_t - Y_{t-1}) = .613\ (X_t - X_{t-1}) + .205(X_{t-1} - X_{t-2}) + 5.6 + \varepsilon_t$

(.853)	(.445)	(10.9)
(.373)	(-.035)	(0.3)

(15) (100) $(Y_t - Y_{t-1}) = .327(Y_{t-1} - Y_{t-2}) + .645(X_t - X_{t-1}) + 1.0 + \varepsilon_t$

(.681)	(.903)	(62.4)
(-.037)	(.387)	(-60.4)

Figure 26. Maximum likelihood estimates and upper and lower confidence limits of the two candidate transfer function models.

	(0,1,0)	(1,0,0)
Autocorrelation of residuals	Q=15.747(on 15 degrees of freedom)	Q=16.58(on 20 degrees of freedom)
Residual sum of squares	$.17830 \times 10^7$	$.16978 \times 10^7$
Significance of parameters	ω_1 slightly insignificant	δ_1 slightly insignificant

Figure 27. Comparison of the (0,1,0) and (1,0,0) transfer functions according to diagnostic indicators

Deciding which of the two models should be selected depends in part on which of the two models has the lower residual sum of squares. Figure 27 shows that model (1,0,0) fits the data slightly better. This is our choice of the transfer function.

Box-Jenkins procedures do not rely solely on the residual sum of squares criterion. Indeed it is not until the later stages of the procedures that this criterion is used. In the identification stage all but two models were eliminated on the basis of the dissimilarity between the theoretical cross-correlation function expected of them and the sample pattern. This basis for eliminating other models was irrespective of their possible low residual sum of squares.

6. Diagnostic Checks

The Box-Jenkins procedures do not stop with the goodness-of-fit criterion. Checks are made to be sure that the residuals of the model chosen in step 5. are not autocorrelated. Were they autocorrelated this would be a warning that the model does not conform to the theoretical pattern expected of it. The residual series of (1,0,0) shown in Figure 28 has nonsignificant autocorrelation as seen in Figure 27.

A second diagnostic check is to verify that there is no cross-correlation between the pre-whitened advertising input series and the residuals. Were there significant cross-correlation this would indicate a lack of independence between the independent variable and the error term, hence confounding the effect of advertising and the unspecified error variable.

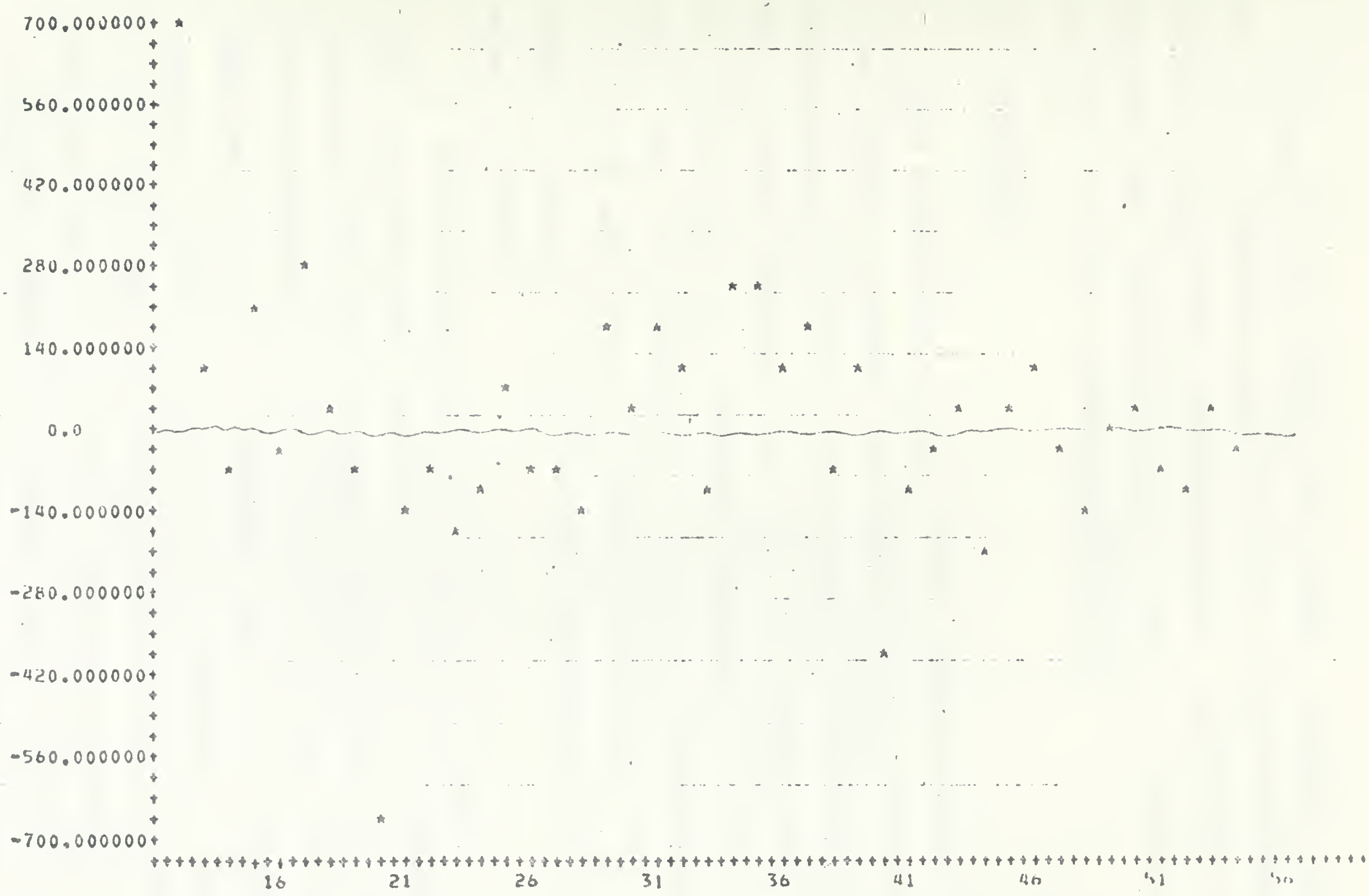


Figure 28. Graph of the residuals of the chosen transfer function model.

Throughout this analysis this concern, a concern shared in theory by all regression model-builders, has been stressed. Figure 29 shows nonsignificant cross-correlation. Should either diagnostic check have been significant, Box and Jenkins 2 describe further procedures to modify the model appropriately.

IV. DISCUSSION

After having gone through the full transfer function analysis of Box-Jenkins, we support the Clarke and McCann conclusion that Lydia Pinkham advertising had no substantial effects past one year. This conclusion would have been even more absolute had we selected the (0,1,0) model which stipulates absolutely no effect past one year. The (1,0,0) model which we marginally preferred over the (0,1,0) has only very small effects past one year. The graph in Figure 30 shows the convergence of both models to the Clarke and McCann estimates.

It should be remembered that there are in general a large number of models which are also quite reasonable representations of the empirical relation between advertising and sales, but which have been eliminated by the procedures outlined in this paper. In comparison to the usual regression "data mining" exhibited in much research -- and also present in Palda's monograph -- one can argue that the Box-Jenkins analysis provides a much more elegant and efficient method for screening a large set of potential models.

Another feature of the present type of analysis is the emphasis it places upon the need for the causal input to be truly

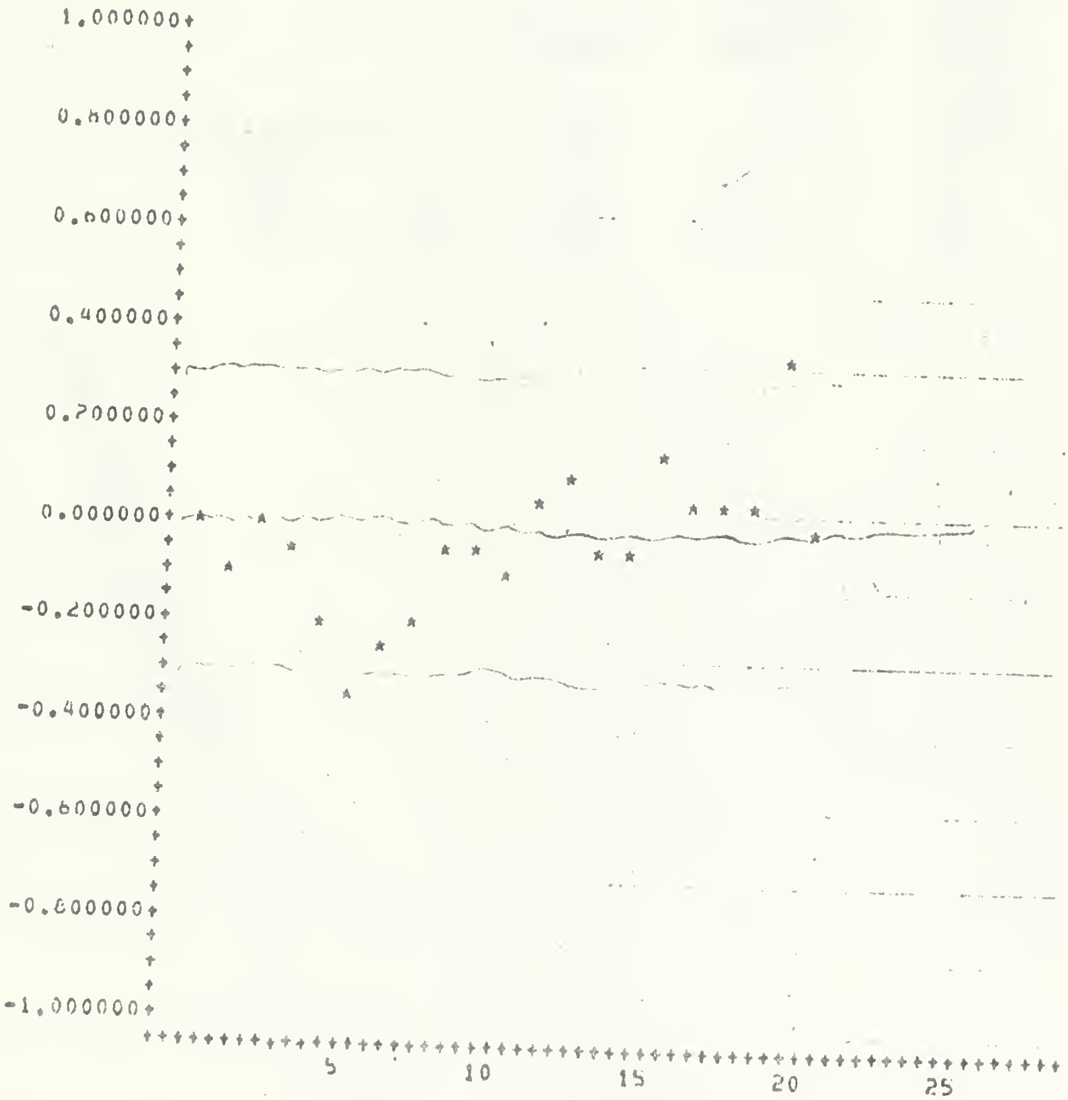


Figure 29. The cross-correlation function between the residual series and the prewhitened advertising input series.

	k	Palda's estimates	Clarke-McCann estimates	Box - Jenkins Maximum likelihood estimates (0,1,0) (14)	Box - Jenkins Maximum likelihood estimates (1,0,0) (15)
	0	.537	.642	.613	.645
Coefficient	1	.337	.297	.205	.211
at	2	.211	.0	.0	.069
lag	3	.133	.0	.0	.007
k	4	.082	.0	.0	.002
	5	.052	.0	.0	.001

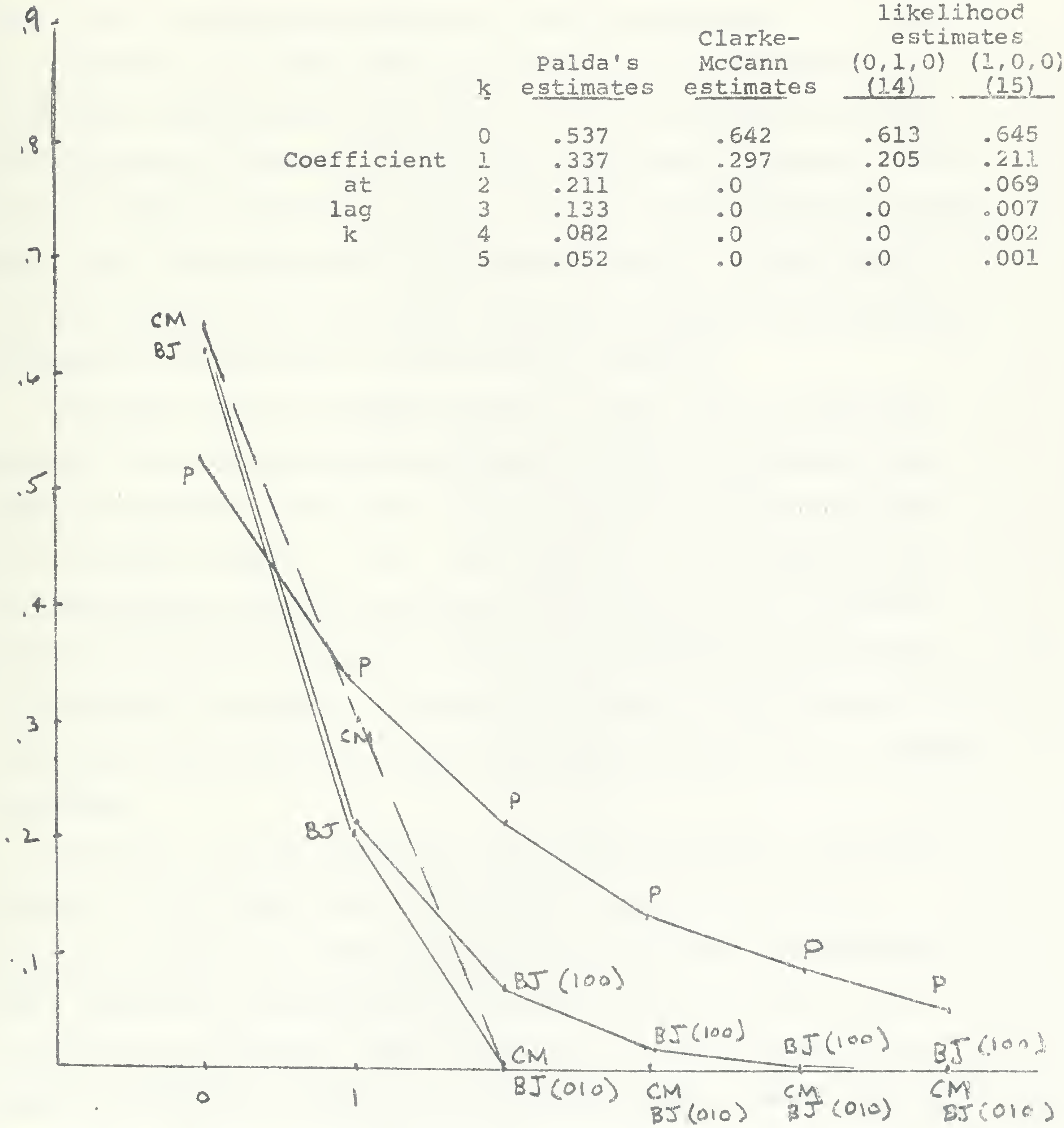


Figure 30. Comparison of Box-Jenkins transfer functions with palda's and Clarke-McCann results. Note Box-Jenkins transfer functions are between first differences of advertising and first differences of sales.

exogenous and the role of experimentation. One might even argue that since the decision maker has a need to understand the responses to his actions, he should actively interfere with the system on a randomized experimental design basis so as to eliminate the possible confounding influences from other unspecified variables. This problem becomes especially acute in the case of advertising studies based, like the present one, on historical data. In such cases the pre-whitening might provide a safeguard against the generation of spurious correlations in the transfer function.

Although the Box-Jenkins analysis is very close to spectral analysis, the mathematics is developed in the more familiar time domain as opposed to the more esoteric frequency domain employed in the latter analysis. Also, even though the terminology used in a Box-Jenkins analysis might seem strange at first, not much effort is required to obtain a rudimentary working knowledge of the procedures involved. In addition, the use of correlation statistics should make for easier adoption than the inverse Fourier transforms.

Overall, the usefulness of a Box-Jenkins approach should be greatest in the cases where a relatively long and continuing history of sales and advertising data is available, and where the sales variable is relatively uninfluenced by other factors controllable by management. The introduction of several input variables is difficult because of computational requirements and the difficulty of pre-whitening several input series independently of each other. This restriction might well prove to be a severe obstacle in many marketing applications. Overcoming this restriction will be the object of further research in building effectiveness models for marketing decision makers.

APPENDIX

STATISTICS USED IN BOX-JENKINS ANALYSIS

For a stationary stochastic process $\{z_t\}$ the autocorrelation function is the set of correlations between z_t and z_{t+k}, $(k=0,1,2,\cdots)$. This correlation, ρ_k, called the autocorrelation at lag k does not depend on t because this is the definition of stationarity. The sample autocorrelation function, r_k, $(k=0,1,2,\cdots)$ is calculated in a way similar to any sample correlation.

$$(16) \qquad r_k = \frac{\frac{1}{n} \sum_{t=1}^{n-k} (z_t-\bar{z})(z_{t+k}-\bar{z})}{\sqrt{\frac{1}{n} \sum_{t=1}^{n} (z_t-\bar{z})^2} \sqrt{\frac{1}{n} \sum_{t=1}^{n} (z_t-\bar{z})^2}}$$

The standard error $SD[r_k]$ has been estimated by Bartlett to be approximately $\sqrt{1/n}$ when k=0 and increasing thereafter. Rather than estimate r_k individually for each lag, it is often convenient, when diagnosing residual error, for example, to estimate whether or not the series, as a whole, has significant autocorrelations. For this we use the Q statistic (developed by Box and Pierce) which is distributed like a Chi-square.

The inversion of a stochastic process means expressing the error as a function of the observed series

$$(17) \qquad \alpha_t = \pi(B) z_t .$$

Assuming that $\pi(B)$ is a k^{th} order equation

$$(18) \qquad \pi_k(B) = 1-\phi_{k_1}B-\phi_{k_2}B^2+\cdots\phi_{k_k}B^k$$

we can get an estimate $\hat{\phi}_{k_k}$, called the sample partial autocorrela-

tion, by solving Yule-Walker equations [2]. If there is an integer p beyond which $\hat{\phi}_{k_k}$ are insignificant, that is, $\phi_{k_k}=0$ for $k=p+1, p+2, \cdots$, we have discovered a parsimonious way of describing the process. The $SD[\hat{\phi}_{k_k}]$ is approximately $\sqrt{1/n}$.

For two stationary stochastic processes $\{\alpha_t\}$ and $\{\beta_t\}$, the <u>cross-correlation function</u> is the set of correlations between α_{t-k} and β_t, $k=0,1,2,\cdots$. The <u>cross-correlation at lag k</u> $\rho_{\alpha\beta}(k)$ is estimated by

$$(19) \qquad r_{\alpha\beta}(k) = \frac{\frac{1}{n}\sum_{t=1}^{n-k} (\alpha_t-\bar{\alpha})(\beta_{t+k}-\bar{\beta})}{SD(\alpha) \quad SD(\beta)} .$$

The standard error of $\hat{r}_{\alpha\beta}(k)$, $SD(\hat{r}_{\alpha\beta}(k))$ is roughly of the order of $\sqrt{1/n}$. Should the sample cross-correlation function be insignificant for all k, one infers that $\{\alpha_t\}$ has no effect on $\{\beta_t\}$.

REFERENCES

1. Aoki, Masanao, Optimization of Stochastic Systems, New York: Academic Press, 1967.

2. Box, George E. P., and Gwilym M. Jenkins, Time Series Analysis, San Francisco: Holden-Day, 1970.

3. Clarke, Darral G., and John M. McCann, "Measuring the Cumulative Effects of Advertising: A Reappraisal," Institute for Research in the Behavioral, Economic and Management Sciences, Krannert Graduate School of Industrial Administration, Purdue University, Paper No. 410, May 1973.

4. Griliches, Zui, "Distributed Lags: A Survey," Econometrica, Vol. 35, No. 1, January 1967.

5. Marschak, Jacob, "On Adaptive Programming," Management Science, Vol. 9, No. 4 (July 1963) pp. 517-26.

6. Nelson, Charles R., Applied Time Series Analysis, San Francisco: Holden-Day, 1973.

7. Palda, K., The Measurement of Cumulative Advertising Effects, Englewood Cliffs: Prentice-Hall, 1964.